

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/29, 15/82, 15/11, 5/10, C07K 14/415, A01H 5/00</b>		A1	(11) International Publication Number: <b>WO 97/23618</b> (43) International Publication Date: <b>3 July 1997 (03.07.97)</b>
(21) International Application Number: <b>PCT/GB96/03191</b> (22) International Filing Date: <b>20 December 1996 (20.12.96)</b>		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: <b>9526218.4 21 December 1995 (21.12.95) GB</b>		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant (for all designated States except US): <b>GENE SHEARS PTY. LIMITED [AU/AU]; Building 5, Suite 1, 105 Delhi Road, North Ryde, NSW 2113 (AU).</b> (72) Inventors; and (75) Inventors/Applicants (for US only): <b>BAUDOT, Gaelle [FR/FR]; 9, rue Jeanne-d'Arc, F-63000 Clermont-Ferrand (FR). GARCIA, Denise [FR/FR]; 2, place Notre-Dame de la Rivière, F-63110 Beaumont (FR). HODGE, Rachel [GB/GB]; 16 Mount Avenue, Leicester LE5 3RN (GB). PEREZ, Pascual [FR/FR]; Chemin de la Pradelle, Varennes, F-63450 Chanonat (FR).</b> (74) Agents: <b>CHAPMAN, Paul, William et al.; Kilburn &amp; Strode, 30 John Street, London WC1N 2DD (GB).</b>			
(54) Title: <b>DNA SEQUENCES CODING FOR A PROTEIN CONFERRING MALE STERILITY</b>			
<p><b>Dendrogram based on the clustal alignment of 41a related sequences</b></p> <p>Percentage amino-acid identity</p>			
(57) Abstract Nucleic acid coding for a protein which confers male sterility is provided together with its use in producing transgenic plants.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## DNA SEQUENCES CODING FOR A PROTEIN CONFERRING MALE STERILITY

5 This invention relates to recombinant, isolated and other synthetic DNA useful in male-sterility systems for plants. In particular, the invention relates to a gene associated with male fertility, labelled Ms41-A, and a recessive mutant form thereof, labelled ms41-A, which confers male sterility. Male-sterile plants are useful for the production of hybrid plants by sexual 10 hybridisation.

15 Hybrid plants have the advantages of higher yield and better disease resistance than their parents, because of heterosis or hybrid vigour. Crop uniformity is another advantage of hybrid plants when the parents are extensively homozygous; this leads to improved crop management. Hybrid seed is therefore commercially important and sells at a premium price.

20 Producing a hybrid plant entails ensuring that the female parent does not self-fertilise. There have been many prior proposals, mechanical, chemical and genetic, for preventing self-pollination. Among the genetic methods is the use of anther-specific genes or their promoters to 25 disrupt the normal production of pollen grains. An anther-specific promoter, for example, can be used to drive a "male-sterility DNA" at the appropriate time and in the right place. Male sterility DNAs include those coding for lytic enzymes, including those that lyse 30 proteins, nucleic acids and carbohydrates. Glucanases are enzymes which break down carbohydrates.

WO-A-9302197 describes recombinant or isolated DNA encoding a glucanase called callase.

Aarts et al, (Nature, 363:715-717 (1993)) have described a gene required for male fertility, isolated from *Arabidopsis*, which has been labelled Ms2.

5 We have now identified and isolated from *Arabidopsis* another gene linked to male fertility. This gene has been labelled Ms41-A. Its mutant, recessive, form is labelled ms41-A and is capable of conferring male sterility. This gene would appear to offer advantages over Ms2 when used  
10 to produce male sterile plants.

Thus, in a first aspect the present invention provides recombinant or isolated Nucleic acid which:

15 a) encodes the Ms41-A protein from *Arabidopsis*;

b) encodes a Ms41-A like protein;

c) encodes the ms41-A protein from *Arabidopsis*;

20 d) encodes a ms41-A like protein;

e) comprises a promoter sequence which regulates expression of the Ms41-A protein from *Arabidopsis* or  
25 a promoter sequence which regulates expression of a Ms41-A like protein; or

f) hybridises under stringent conditions to Nucleic acid a), b), c), d) or e) or would do so but for the  
30 degeneracy of the genetic code.

In one embodiment of a) above, the Nucleic acid encodes a protein having an amino acid sequence as shown in figure 4. Although figure 4 relates only to a protein of

Arabidopsis, those skilled in the art will readily be able to identify equivalent proteins from other members of the family Brassicaceae or indeed similar proteins from other commercially important plant families, ie 5 Ms41-A like proteins.

In turn the equivalent genes may be identified by hybridisation studies, restriction fragment length polymorphism (RFLP), degenerate PCR and other methods 10 known in the art. Genes or other DNA sequences, whether natural, engineered or synthetic, encoding closely equivalent proteins may for example hybridise under stringent conditions (such as at approximately 35°C to 15 65°C in a salt solution of approximately 0.9 molar) to the *Arabidopsis* gene, or fragments of it of, for example, 10, 20, 50 or 100 nucleotides. A 15-20 nucleotide probe would be appropriate under many circumstances.

In the context of the present invention, "Nucleic acid 20 which encodes" includes all nucleic acid, eg DNA sequences which will, when expressed, give rise to the protein. Examples of such DNA sequences include, but are not limited to, ones which comprise non-coding regions, 25 e.g introns, sequences which include leader sequences and/or signal sequences, or simply comprise a coding sequence for the protein. The skilled person will also appreciate that, due to codon degeneracy, there will, for example, be a number of DNA sequences capable of coding for the Ms41-A protein or a Ms41-A like protein.

30 In general, the Nucleic acid of the invention will comprise at least a direct coding sequence for the protein as well as a promoter and transcription termination sequence. The promoter can itself comprise

5 only those sequences, or elements, necessary for the correct initiation of transcription (which regions can be described as transcription initiation regions, for instance), or, alternatively, it can include regions of sequence which are not directly involved in the initiation of transcription, i.e. a complete promoter can be employed.

10 A preferred coding sequence described in this specification is from *Arabidopsis* and can be isolated by methods known in the art, for example by (a) synthesising cDNA from mRNA isolated from *Arabidopsis*, (b) isolating this cDNA. This cDNA can, in turn, be used (c) as a probe to identify regions of the plant genome of a chosen 15 member of another plant species, eg Maize, that encode mRNA of interest and (d) identifying the upstream (5') regulatory regions that contain the promoter of this DNA.

20 A particularly preferred DNA sequence is that shown in figure 3, and more particularly, the sequence shown in figure 3 which commences with the base pair labelled 1, as will subsequently be described in the examples. Those skilled in the art will, with the information given in this specification, be able to identify with sufficient 25 precision the coding regions and to isolate and/or recombine DNA containing them.

30 The Nucleic acid of the invention can be used to confer male sterility on plants. For instance, the recessive form of the gene, ie ms41-A can be used to transform a plant. Alternatively, the dominant form, ie Ms41-A can be downregulated in some way.

As discussed herein, the Nucleic acid can include a

promoter, and to increase the likelihood of male sterility being conferred it is possible to use promoters which drive expression in particular plant tissues which are involved in the control of fertility. Examples of such promoters are those which are tapetum-specific, for example a Brassicaceae A3 or A9 promoter, described in WO-A-9211379, and the A6 promoter described in WO-A-9302197. Both WO-A-9211379 and WO-A-9302197 are hereby incorporated by reference.

10 Because of the natural specificity of the regulation of expression of the Ms41-A or Ms41-A like gene, it is not necessary for the Ms41-A promoter to be linked to specific disrupter DNA to provide a useful male-sterility 15 system (although it can be); non-specific disrupter DNA can be used.

20 Ms41-A like promoters from other plant species, eg from Maize, and modified Ms41-A promoters can be used, and if necessary located or identified and isolated as described above for the Ms41-A coding sequences, *mutatis mutandis*.

25 Ms41-A or Ms41-A like promoter-containing DNA in accordance with the invention can, as indicated above, be used to confer male sterility on plants, particularly those belonging to the family *Brassicaceae*, in a variety of ways as will be discussed below. In an important embodiment of the invention, therefore, a promoter as described above is operatively linked to DNA which, when 30 expressed, causes male sterility.

Since an effective sterility system is complete, propagation of the seed parent must proceed either by asexual means or via the pollination of the male-sterile

by an isogenic male-fertile line, and the subsequent identification or selection of male sterile plants among the offspring. Where vegetative propagation is practical, the present invention forms a complete system for hybrid production. Where fertility restoration is necessary to produce a seed crop, the present invention forms the basis of a new male sterility system. In some seed crops where the level of cross pollination is high, seed mixtures may enable restoration to be bypassed. The male sterility will be particularly useful in crops where restoration of fertility is not required, such as in the vegetable *Brassica* spp., and such other edible plants as lettuce, spinach, and onions.

Nucleic acid in accordance with the invention and incorporating the Ms41-A or Ms41-A like promoter can drive male sterility DNA thereby producing male sterile plants, which can be used in hybrid production.

A construct comprising a promoter operatively linked to a male sterility DNA can be transformed into plants (particularly those of the genus *Brassica*, but also other genera such as *Nicotiana* and *Hordeum*) by methods which may be well known in themselves. This transformation results in the production of plants, the cells of which contain a foreign chimeric DNA sequence composed of the promoter and a male sterility DNA. Male-sterility DNA encodes an RNA, protein or polypeptide which, when produced or over-produced in a stamen cell of the plant, prevents the normal development of the stamen cell. The Ms41-A or Ms41-A like promoter may be used to drive a variety of male sterility DNA sequences which code for RNAs, proteins or polypeptides which bring about the failure of mechanisms to produce viable male gametes. The

invention is not limited by the sequence driven, but a number of classes and particular examples of male sterility promoter-drivable sequences are preferred.

5 For example, the drivable male sterility DNA may encode a lytic enzyme. The lytic enzyme may cause degradation of one or more biologically important molecules, such as macromolecules including nucleic acid, protein (or glycoprotein), carbohydrate and (in some circumstances) 10 lipid.

Ribonuclease (such as RNase T1 and barnase) are examples of enzymes which cause lysis of RNA. Examples of enzymes which lyse DNA include exonucleases and endonucleases, 15 whether site-specific such as EcoRI or non-site-specific.

Actininidin is an example of a protease, DNA coding for which can be suitable male sterility DNA. Other examples include papain zymogen and papain active protein.

20 Lipases whose corresponding nucleic acids may be useful as male sterility DNAs include phospholipase A<sub>1</sub>.

Male sterility DNA does not have to encode a lytic 25 enzyme. Other examples of male sterility DNA encode enzymes which catalyse the synthesis of phytohormones, such as isopentyl transferase, which is involved in cytokinin synthesis, and one or more of the enzymes involved in the synthesis of auxin. DNA coding for a 30 lipoxygenase or other enzymes having a deleterious effect may also be used.

As mentioned above, one way to confer male sterility will be to downregulate the Ms41-A or Ms41-A like gene. This

could be achieved by the use of antisense DNA. Introducing the coding region of a gene in the reverse orientation to that found in nature can result in the down-regulation of the gene and hence the production of less or none of the gene product. The RNA transcribed from antisense DNA is capable of binding to, and destroying the function of, a sense RNA version of the sequence normally found in the cell thereby disrupting function.

10

It is not crucial for antisense DNA solely to be transcribed at the time when the natural sense transcription product is being produced. Antisense RNA will in general only bind with its sense complementary strand, and so will only have its toxic effect when the sense RNA is transcribed. Antisense DNA corresponding to some or all of the DNA encoding the Ms41-A or Ms41-A like gene product may therefore be produced not only while the gene is being expressed. Such antisense DNA may be expressed constitutively, under the control of any appropriate promoter.

25

It is also the case that one may wish to restore male fertility in later generations. this can also be achieved using antisense nucleic acid, eg nucleic acid which is antisense for a DNA molecule encoding ms41-A.

30

Thus, in a second aspect, the present invention provides Antisense nucleic acid which includes a transcribable strand of DNA complementary to at least a part of a DNA molecule of the invention.

In one embodiment of this aspect the antisense nucleic acid is under the control of a constitutive promoter,

such as the CaMV35S promoter.

5 A still further example of male sterility DNA encodes an RNA enzyme (known as a ribozyme) capable of highly specific cleavage against a given target sequence (Haseloff and Gerlach *Nature* 334 585-591 (1988)). Like antisense DNA, ribozyme DNA (coding in this instance for a ribozyme which is targeted against the RNA encoded by the Ms41-A or Ms41-A like gene) does not have to be expressed only at the time of expression of the Ms41-A or 10 Ms41-A like gene. Again, it may be possible to use any appropriate promoter to drive ribozyme-encoding DNA, including one which is adapted for constitutive expression.

15 According to a further aspect of the invention, there is therefore provided DNA encoding a ribozyme capable of specific cleavage of RNA encoded by a DNA molecule of the invention. Such ribozyme-encoding DNA would be useful in 20 conferring male sterility on members of, eg the family *Brassicaceae*.

In addition, there are other useful methods which can be employed for the downregulation of the Ms41-A or Ms41-A 25 like DNA sequences. Some examples of these are as follows:

i) expression of an antibody or antibodies, domains or fragments thereof against the Ms41-A or 30 a Ms41-A like protein;

ii) expression of mutant versions of the Ms41-A or of a Ms41-A like protein which may interfere with the function of the normal protein;

iii) by creation of mutations in the Ms41-A sequence or the the Ms41-a like sequence with the result that mutant plants can be used in the recessive AMS system as hereinbefore described; and

5

iv) expression of mRNA binding proteins that will interfere specifically with Ms41-A or Ms41-A like transcription.

10 In preferred embodiments of DNA sequences of this invention 3' transcription regulation signals, including a polyadenylation signal, may be provided. Preferred 3' transcription regulation signals are derived from the Cauliflower Mosaic Virus 35S gene. It should be 15 recognised that other 3' transcription regulation signals could also be used.

20 Recombinant DNA in accordance with the invention may be in the form of a vector. The vector may for example be a plasmid, cosmid or phage. Vectors will frequently include one or more selectable markers to enable selection of cells transfected (or transformed: the terms are used interchangeably in this specification) with them and, preferably, to enable selection of cells harbouring 25 vectors incorporating heterologous DNA. Appropriate start and stop signals will generally be present. Additionally, if the vector is intended for expression, sufficient regulatory sequences to drive expression will be present; however, DNA in accordance with the invention 30 will generally be expressed in plant cells, and so microbial host expression would not be among the primary objectives of the invention, although it is not ruled out. Vectors not including regulatory sequences are useful as cloning vectors.

Cloning vectors can be introduced into *E. coli* or another suitable host which facilitate their manipulation. According to another aspect of the invention, there is therefore provided a host cell transfected or transformed 5 with DNA as described above.

DNA in accordance with the invention can be prepared by any convenient method involving coupling together successive nucleotides, and/or ligating oligo- and/or 10 poly-nucleotides, including *in vitro* processes, but recombinant DNA technology forms the method of choice.

Ultimately, DNA in accordance with the invention (whether 15 (i) Ms41-A gene, ms41-A gene, Ms41-A like gene or ms41-A like gene (ii) antisense DNA to any option listed in i), ribozyme DNA targeted to RNA for any option listed in i) or DNA comprising a promoter as described herein used to drive expression of a disrupter sequence, eg encoding Barnase) will be introduced into plant cells, by any 20 suitable means.

According to a further aspect of the invention, there is provided a plant cell including DNA in accordance with the invention as described above.

25 Preferably, DNA is transformed into plant cells using a disarmed Ti-plasmid vector and carried by *Agrobacterium* by procedures known in the art, for example as described in EP-A-0116718 and EP-A-0270822. Alternatively, the 30 foreign DNA could be introduced directly into plant cells using an electrical discharge apparatus. This method is preferred where *Agrobacterium* is ineffective, for example where the recipient plant is monocotyledenous. Any other method that provides for the stable incorporation of the

DNA within the nuclear DNA of any plant cell of any species would also be suitable. This includes species of plant which are not currently capable of genetic transformation.

5

Preferably DNA in accordance with the invention also contains a second chimeric gene (a "marker" gene) that enables a transformed plant containing the foreign DNA to be easily distinguished from other plants that do not contain the foreign DNA. Examples of such a marker gene include antibiotic resistance (Herrera-Estrella et al., *EMBO J.* 2, 987-995 (1983)), herbicide resistance (EP-A-0242246) and glucuronidase (GUS) expression (EP-A-0344029). Expression of the marker gene is preferably controlled by a second promoter which allows expression in cells other than the tapetum, thus allowing selection of cells or tissue containing the marker at any stage of regeneration of the plant. The preferred second promoter is derived from the gene which encodes the 35S subunit of Cauliflower Mosaic Virus (CaMV) coat protein. However any other suitable second promoter could be used.

25

A whole plant can be regenerated from a single transformed plant cell, and the invention therefore provides transgenic plants (or parts of them, such as propagating material) including DNA in accordance with the invention as described above. The regeneration can proceed by known methods. When the transformed plant flowers it can be seen to be male sterile by the inability to produce viable pollen. Where pollen is produced it can be confirmed to be non-viable by the inability to effect seed set on a recipient plant.

30

Preferred features of each aspect of the invention are as

for each other aspect *mutatis mutandis*.

The invention will now be illustrated by a number of non-limiting examples, which refer to the accompanying 5 drawings, in which:

FIGURE 1: shows a Southern Blot of *Hind*III-cut genomic DNA from 21 *ms41-A* plants demonstrating linkage of the 35S-Ac element to *ms41-A*;

10 FIGURE 2: shows a schematic diagram of the region containing the *MS41-A* locus cloned in lambda MSE3. The position of insertion of the 35S-Ac is indicated; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; S, *Sae*I;

15 FIGURE 3: shows the genomic DNA sequence of the *MS41-A* gene. The sequence is numbered from the putative transcriptional start point of the 20 *MS41-A* message. The predicted amino-acid sequence of *MS41-A* is shown together with the restriction sites;

25 FIGURE 4: shows the predicted amino acid sequence of *MS41-A*;

FIGURE 5: shows the oligonucleotides used to examine excision events of 35S-Ac from the *ms41-A* locus;

30 FIGURE 6: shows DNA sequences left by 35S-Ac excision events at the *ms41-A* locus;

FIGURE 7: shows a diagram of the *MS41-A*

promoter-GUS and MS41-A promoter-Barnase chimeric genes;

5 FIGURE 8: shows a diagram of the MS41-A promoter-antisense MS41-A and CaMV 35S promoter-antisense and sense MS41-A chimeric genes;

10 FIGURE 9: shows sequence alignments of proteins related to MS41-A;

FIGURE 10: shows a partial DNA sequence and predicted amino acid translation of Zm41-A;

15 FIGURE 11: shows a dendrogram of MS41-A related sequences;

20 FIGURE 12: shows the nucleotide sequence of the Z31 Zm41-A gene. The portion of the sequence corresponding to putative coding region is shown in bold type capital letters. ♦ indicates putative first methionine deduced in frame with cDNA Zm41-A and 5'RACE products. \* indicates the start of the longest 5'RACE 25 product. ▽ indicates the start of Zm41-A cDNA. 12 exons are present and the translation is stopped in exon 11, the stop codon is TGA (□). Non spliced DNA present in some RACE products is underlined;

30 FIGURE 13: shows restriction maps of Z31, Z33 and Z35 genomic clones isolated with cDNA of Zm41-A. EI, HIII, NI and SI indicate restriction sites of endonucleases EcoRI, HindIII, NcoI and

*Sall*, respectively. \* indicates the start of the longest RACE product. ▼ indicates the start of *Zm41-A* cDNA. Dotted lines indicate homologous regions and ▲ indicates deletions;

5

FIGURE 14: shows clustal V alignment between the protein deduced from the *Zm41-A* cDNA and from the genomic longest open reading frame of *Z31*;

10

FIGURE 15: shows the nucleotide sequence of the *Z33 Zm41-A* gene. The portion of the sequence corresponding to DNA transcription is shown in bold type capital letters. Non spliced DNA present in some RACE products is underlined. This gene is truncated and only exons 3,5 and 6 are present; and

20 FIGURE 16: shows the nucleotide sequence of the *Z35 Zm41-A* gene. The portion of the sequence corresponding to DNA transcription is shown in bold type capital letters. Non spliced DNA present in some RACE products is underlined. This gene is truncated and only exons 3,4,5 and 6 are present.

25

#### Example 1

30 Isolation of a gene required for male fertility in *Arabidopsis thaliana*

i) Isolation and phenotype of the *ms41-A* male sterile mutant.

The method used to identify a gene required for male fertility in *Arabidopsis thaliana* was transposon tagging. This method is a powerful technique for isolating genes which encode unknown products, allowing genes identified only by their mutant phenotype, to be cloned. 5 *Arabidopsis thaliana* is a widely used model species that is an ideal plant for transposon tagging of genes, since it is a transformable diploid with a very small genome. Thus the chance of tagging desired genes is maximised. 10 Additionally *Arabidopsis* is a *Brassicaceae* and is thus very closely related to important crop plants such as *Brassica napus* (Oil Seed Rape).

Transposon tagging was achieved by transformation of C24 15 *Arabidopsis* roots with modified autonomous Ac elements from Maize: D Ac and 35S Ac inserted into the leader of the GUS reporter gene in the reverse orientation (Constructs described in Finnegan et al., *Plant Molecular Biology*, 22:625-633 (1993)). (As this work was in progress 20 the first reports of gene tagging with similar Ac elements in heterologous plant species were published; a pH controlling gene from Petunia: Chuck et al., *Plant Cell*, 5:371-378 (1993)); the *Arabidopsis* DRL1 locus: Bancroft et al., *Plant Cell*, 5:631-638 (1993)) and the 25 *Arabidopsis* Albino gene (Long et al., *Proceedings of the National Academy of Sciences U.S.A.*, 90:10370-10374 (1993)).

Transformed plants were regenerated and the T2 progeny 30 analysed for GUS activity and by molecular analysis. This demonstrated that the 35S Ac transposed quite efficiently (in 30% to 40% of progeny). The T3 progeny families derived from 279 selected T1 plants were then visually screened for mutants affected in male sterility.

A few fertility-reduced or sterile plants were recovered, some possessing additional abnormalities. A male sterile mutant (ms41-A) which appeared in family 41 had collapsed anthers with empty locules. Only one sterile plant was recovered from more than 2000 T3 siblings in this family. After cross-pollination with wild type pollen, elongation of siliques was observed, confirming that female fertility is unaffected by the mutation.

From the above cross 21 F1 individuals were grown and allowed to self pollinate to produce F2 seed ; all the F1 plants were completely fertile suggesting that the mutation is recessive. The first analysis of 6 different F2 populations confirmed the recessive character of the mutation, as male sterility reappeared in a small proportion of each F2 population, with all other siblings presenting a wild type phenotype. Moreover, the vegetative development of the male sterile plants was identical to wild type *Arabidopsis*. The observed frequency of male transmission of the mutation suggests a non-classical mendelian inheritance for a single recessive mutation - the frequencies of mutant plants in the F2 populations were: 16.8 ; 13.0 ; 11.9 ; 12.7 ; 15.4 and 17.0 %. The expected frequency of mutant plants is 25 % or a 3 to 1 ratio of wild type to mutant plants. In this case there is a ratio of approximately 7 to 1 wild type to mutant plants. A homogeneity test on the data of the 6 F2 populations presented concludes that there is homogenous transmission of the male sterile phenotype (Chi square with 5 degrees of freedom = 8.69,  $0.10 < P < 0.20$ ).

Proof of reduced transmission of Ms41-A through the male

gametophyte was obtained by genetic mapping of Ms41-A. The hypothesis was that markers genetically linked to Ms41-A but present on the homologous chromosome (in repulsion) on a F1 cross with an Ms41-A plant should be 5 over-represented in the derived F2 population. The F1 crosses were made with 5 tester lines, one for each chromosome, constructed by Marteen Korneef (described in; O'Brian S.T. (ed) Genetic maps of complex genomes, Book 6, Plant Cold Spring Harbor Laboratory Press, Cold Spring 10 Harbor, New York, pp 94-97 (1990)), and linkage of Ms41-A was demonstrated with markers on the lower part of chromosome 1. Compiled recombination data of 2 populations (476 and 540 individuals) were analysed by 15 the Map Maker software version 2 (Lander et al., *Genetics*, 121:174-181 (1987))).

Ms41-A is between apetala 1 (8.1 cM) and glabra 2 (9.8 cM) and 40.2 cM away from than chlorina 1. In the first 20 F2 population, the deficit of Ms41-A plants was observed as before (14.7% of plants were male sterile) and it was correlated with the expected increase of apetala 1 and glabra 2 plants (29 % and 31.5 % respectively) ; the most distal marker, chlorina 1 behaves quite normally (22.3 %). In the second F2, where the penetrance of the Ms41-A 25 is less affected (18.3 %), the over representation is not as prevalent (as expected) ; only the proportion of glabra 2 plants appears to be slightly increased (27.2 %).

30 Microscopic observations of microsporogenesis in the male sterile Ms41-A plants revealed that the tetrads release abnormal microspores which degenerate rapidly. By aniline blue staining the tetrads appear abnormal with irregular shaped cells and with great variation in cell size. Moreover there is a mixed population of meiocytes, dyads

(a stage not usually observed in *Arabidopsis*) and tetrads in the same anther. The defect apparently lies just before or during meiosis. Cytological observations on fixed young anther buds reinforce this finding, since at 5 meiosis the meiocytes are affected but the tapetum behaves normally. No differences were observed cytologically between the Ms41-A heterozygote and wild type plants.

10 One other gene required for male-fertility (also in *Arabidopsis*) has been described previously (Aarts et al., *Nature*, 363:715-717 (1993)). Plants with a mutation in this gene (Ms2) were grown together with Ms41-A plants. In certain conditions, especially after the 15 plants had been flowering for a long time the ms2 but not the Ms41-A plants reverted to male fertility.

20 ii) Linkage of a transposed 35S Ac with the mutant phenotype

To determine if the Ms41-A mutation was due to the insertion of a 35S-Ac element, HindIII-cut DNA from five Ms41-A F1 individuals was analysed by Southern blotting using a 5'Ac fragment (2.5 Kb EcoR I fragment from 25 pBGS335RI (Finnegan et al., *Plant Molecular Biology*, 22: 625-633 (1993)) as a probe. Two identical Ac bands were present in the five mutant plants :

30 - the internal Ac Hind III 1.6 kb band and  
- a junction 3' Ac band of approximately 2.8 kb, which differs from the expected non-transposed 35S Ac (2.1 kb).

This indicates the presence of only one 35S Ac element which has transposed in the parental male sterile plant,

or more likely in its parents. To determine linkage between this 35S Ac element and the Ms41-A phenotype, 24 Ms41-A plants from each of 6 different F2 populations were analysed by PCR for the presence of the Ac element 5 using oligonucleotides:-

5' H (5' AAGGATCCTGGCAAAGACATAAATC 3') and  
Ac12 (5' AGATGCTGCTACCCAATCTTTGTGC 3').

The results were as follows :

10	F2 41-A-A	23 positives out of 24
	F2 41-A-B	5 "
	F2 41-A-C	23 "
	F2 41-A-D	10 "
	F2 41-A-E	24 "
	F2 41-A-F	3 "

15 If the Ac element is linked to Ms41-A all male sterile plants should have the Ac element, however if the Ac is not linked only 3/4 of Ms41-A plants should have the Ac element. The results obtained indicate complete linkage 20 only in the 41-A-E population. The lack of linkage in the other populations may be due to frequent imprecise excision of the Ac element from the Ms41-A locus leaving a mutation in Ms41-A.

25 To confirm linkage, the most stable population, 41-A-E, was analysed by Southern blotting with a probe that contained both a region of the transposed Ac element and 3' flanking plant DNA. To generate this probe DNA from a Ms41-A plant was digested with SspI, religated and 30 amplified by PCR using Ac oligonucleotides:-

Ac 11 (5' CGTATCGGTTTCGATTACCGTATT 3') and  
Ac 12 (5' AGATGCTGCTACCCAATCTTTGTGC 3').

The 1.1kb inverse PCR (IPCR) fragment generated contained 500 bp of Ac and the remainder consisted of 3' flanking

*Arabidopsis* DNA.

DNA from plants of the F2 population 41-A-E was digested with HindIII and probed with the 3' IPCR fragment. 21 new 5 F2 mutant individuals and 28 male fertile F2 plants were analysed, the selfed progenies of the latter were checked for the presence of mutant plants revealing that 15 of the 28 were heterozygous for Ms41-A. All of the 21 mutant plants (Figure 1) and those heterozygotes segregating the 10 mutation in the F3 showed the same transposed 35S Ac revealed by the 2.8 kb specific band and the Ac internal 1.6 kb band. A 3.3 kb band, corresponding to the wild 15 type allele is detectable in most of the F2 mutants; this is probably due to somatic excision of Ac and confirms that the transposed Ac element is still active. These results confirm that the 35S Ac is located in or in the vicinity of the Ms41-A gene.

## iii) Genomic clones and cDNAs of the Ms41-A gene

20 Two different genomic libraries - one MboI partial library in EMBL 3A ( Clontech) and one HindIII partial in Lambda Dash II ( T. Pelissier, S.Tutois and G. Picard, unpublished) were screened with the 3' IPCR cloned 25 product. Four different clones spanning the mutated region, were characterised by Southern analysis. One of them, lambda MSE3, which spans the transposon insertion site, was used for fine mapping. It contains the IPCR hybridising fragments detected on a genomic Southern 30 (HindIII 3.3 kb, SspI 1.8 kb and PstI 4 kb). The entire plant DNA insert in MSE3 is contained on 4 SalI fragments; S1 (5kb), S2 ( 4.9kb), S3 (4.3kb) and S4 (2.3kb) (Figure 2). The S3 fragment contains the plant DNA from the IPCR product.

After sequencing the IPCR product to determine the plant sequence 3' of the Ac element, more than 5000 bp of genomic sequence was obtained from MSE3 (3100 bp from the 5' Ac flanking region and 1900 bp at the 3'). The 5 genomic sequence is presented in figure 3 and is indexed according to the putative transcription initiation site determined by 5' RACE (see below). One of the SalI sites of the fragment S3 is positioned at 2061 bp the other one is situated 5' upstream an EcoRI site (-1753 bp) and has 10 not been sequenced. The transposon is inserted at position +318 bp.

To identify mRNAs expressed in the region of the transposon insertion site, three *Arabidopsis* cDNA 15 libraries were probed with either the S1 or S3 fragments; a developing flower buds library (young buds) (Weigel et al., *Cell*, 69:843-859 (1992)), a library from flowers at late stages (after stage 10) (Hofte et al., *Plant Journal*, 4:1051-1061 (1993)) and an immature siliques 20 library (Giraudat et al., *Plant Cell*, 4:1251-1261 (1992)).

Two classes of cDNAs were recovered with the S3 fragment 25 as a probe and characterised.  
- a 1.9 kb cDNA (W11), isolated from the developing flower buds library. Its 3' end is located 1.5 kb upstream of the 3' 35S Ac end, suggesting that it is not linked to the Ms41-A phenotype. Sequencing of the extremities revealed that the EcoRI site (-1753bp in 30 figure 3) is present in the 3' part of this mRNA.

- a 0.8 kb cDNA (G6), isolated from the immature siliques library but also present in the developing flower buds library. Comparison of G6 and genomic sequences shows

that the transposon insertion site is 1440 bp upstream of the 5' end of the longest G6 cDNA (861 bp). In addition, the lack of a methionine codon in the 5' sequence of G6 indicated that this cDNA was not full-length. Further attempts at obtaining longer cDNAs from the three libraries were unsuccessful.

Another cDNA (A6) of approximately 1Kb was isolated using the S1 fragment as a probe. It maps downstream of the G6 message.

Out of the 3 transcription units in the vicinity of the transposon insertion site, the best candidate for the Ms41-A mRNA was that corresponding to G6. To obtain a full-length G6 cDNA, primers were designed to the 5' end of the longest G6 cDNA and used in a 5' RACE reaction (5' AmpliFinder kit, Clontech). This proved unsuccessful, probably due to the 5' end of G6 lying far upstream of the longest cDNA obtained. Therefore primers were designed to regions of the genomic sequence that were upstream of the 5' end of the longest G6 cDNA. These, in combination with primers designed to the G6 cDNA, were used in RT-PCR reactions to define the extent of the G6 transcribed region. Results obtained suggested that the G6 message was at least 1 kb longer than the longest G6 cDNA obtained, and that the upstream sequence contained an intron of about 450 bp.

The G6 transcriptional start site was finally mapped by 5' RACE using primers Z3 (5' TTATCATCAACATGCCATCGAATCTGCCG 3', positions 494-464 bp in Figure 3); and W1 (5' AAAGTAGTAAACCCTAGAG 3', positions 279-260 bp). RT-PCR was then used to recover a nearly full-length G6

message. Comparison of the G6 and genomic sequences shows that the first ATG is situated at position 157 bp; thus G6 putatively encodes a protein of 584 amino acids (Figure 4). Over the region of overlap the cDNA and genomic DNA sequences were identical. This deduced protein has no significant homology to proteins of known function on the Genebank, EMBL and NBRF databases. The coding sequence consists of three exons, the first of which has been disrupted by the insertion of the 35 Ac element at amino acid position 54 in the Ms41-A mutant. This is strong evidence that G6 corresponds to Ms41-A. Final confirmation was obtained by analysis of phenotypes and DNA sequences around the Ac insertion site in Ms41-A progeny plants in which the 35S Ac element has excised.

To induce somatic excision of the 35S Ac element, plants were regenerated from liquid root cultures from single individuals derived from two different test-crosses. These crosses were between plants (A and B) that had only one Ac element but were still male sterile due to imprecise excision of the other Ac element, and male fertile plants that were heterozygous for Ms41-A: 35S Ac. This material was chosen because of the higher percentage of male sterile plants (40% instead of 20%, 50% instead of 25%?) than in a normal F2 population. Regenerants from clones representing male sterile plants were scored for male fertility. Numerous completely fertile plants were obtained from some individuals, however from 5 different regenerated plants from 4 different individuals, 7 different "revertant siliques" were obtained.

DNA from revertant plants or from progeny from "revertant siliques" was analysed by PCR for excision of the Ac element and PCR products cloned to determine the sequence

left by the Ac element (footprint). The oligonucleotides presented in Figure 5 were used : Ac 11 with W2 for the presence of the 3' junction, Ac 14 with G6 5'-11 for the 5' junction and W2 with G6 5'-11 or with Z3 for the 5  
excision allele(s). The PCR fragments derived from W2 with G6 5'-11 or with Z3 were cloned in the pGEM-T vector (Promega) and sequenced for all revertants. Previously 10 junction products were sequenced confirming the presence of the typical target duplicated sequence of 8 base pairs : CTCCTCTC (positions 311 to 318 in Figure 3).

The genotypes of 7 revertant plants or sectors were determined and are presented in figure 5. For all of them 15 an allele restoring the open reading frame is observed which is the same as the wild type in 4 cases, a 3 bp insertion in 2 cases and a 6 bp insertion in one case. Footprints destroying the coding phase are observed in different revertants and also in the female parents (2 different 7 bp insertions and 2 different 5 bp insertion, 20 and one with the addition of a 9 bp insertion which also introduces an in frame, TGA, stop codon). Their presence is always associated with segregation of male sterile individuals in the progeny. These results demonstrate that the Ms41-A protein has a determinant role in male 25 fertility and that the Ms41-A gene has been tagged with the 35S Ac element.

iv) Ms41-A genetic mapping

30 Classical genetic mapping of Ms41-A with visual phenotypic markers has been described previously in section i) of this example. It places the Ms41-A locus near the bottom of chromosome 1. To determine if the Ms41-A mutation has been isolated previously in

Arabidopsis the mutation was mapped more precisely using recombinant inbred lines made by Caroline Dean (Lister et al., *Plant Journal*, 4:745-750 (1993)). This method requires the identification of restriction enzyme 5 fragment length polymorphisms (RFLPs) between the two parental lines (Columbia and Landsburg erecta) which are in, or near the Ms41-A locus. Polymorphisms were not found in Ms41-A or 5' of it, however the downstream cDNA, 6A, gives a HhaI polymorphism. Results, processed 10 by MapMaker version 2, have positioned Ms41-A near the marker m532 (1.3 cM) and marker g17311 (4.6 cM). Those RFLP markers are situated on chromosome 1 close to the ADH locus, and map in the vicinity of glabrous 2 and apetala 1 on the integrated Arabidopsis genetic map 15 (Hauge et al., *Plant Journal*, 3:745-754 (1993)).

Ms41-A is a new male-sterile mutant. It is not allelic to ms1 (Van der Veen and Wirtz, *Euphytica*, 17: 371-XXX (1968)) ms3, ms5, ms10, ms11 or ms12 (Chaudhury 1993). 20 It is also different to the Ms2 gene (Aarts et al., *supra*).

v) Abundance of the Ms41-A message

25 Ms41-A is expressed in 7 day old seedlings, in young floral buds and in immature siliques (cDNA libraries and RT-PCR data). The mRNA could not be detected in these tissues by Northern blotting using poly A+ mRNA which had been used successfully in RT-PCR analysis for the Ms41-A 30 message. Thus the Ms41-A message appears to be of very low abundance; approximately 10 fold lower than another message required for male fertility in *Arabidopsis*, Ms2, in the same cDNA library (1 out of 12 000 plaques for Ms2 (Aarts et al., *supra*) versus 1 out of 125 000 for

Ms41-A) .

Example 2

5 Isolation of the Ms41-A promoter and fusion to the  
 $\beta$ -Glucuronidase (GUS) reporter gene

To attempt to determine the extent of utility of the  
10 Ms41-A promoter in male sterility systems putative Ms41-A  
promoter fragments were linked to the reporter gene GUS  
and transformed into *Arabidopsis* and tobacco. This will  
reveal more precisely the spatial and temporal expression  
patterns of the Ms41-A gene and determine whether the  
15 low abundance of the Ms41-A transcript is due to weak  
expression or transcript instability.

Two promoter fragments, -903 (Hind III) to +79 (Short  
10 promoter) and -1753 (EcoR I) to +79 (Long promoter), have  
been fused to the GUS gene (transcriptional fusions) to  
20 produce the binary vectors pBIOS 176 and pBIOS 177  
(Figure 7).

These plasmids were constructed as follows:-  
The primers Y7 (positions -1799 to -1782 in Figure 3)  
25 5' CCTAACTTCTTGCGGGC 3'  
and W3 Xba (positions 84 to 59 in Figure 3)  
5' GATCTAGACCGTGATGTCTTAGAAGG 3'  
were used in a PCR to recover a 1883 bp Ms41-A promoter  
fragment. This was cloned into the vector pGEM-T  
30 (Promega) forming pS11. This plasmid was introduced into  
a *dam*, *dcm* minus *E.coli* strain (SCS 110) thus allowing  
the XbaI restriction enzyme to cleave the XbaI site. The  
985 bp HindIII, XbaI fragment of pS11 was cloned between  
the HindIII and XbaI sites of pBI121 (replacing the 35S

CaMV promoter of this plasmid) forming plasmid pBIOS176. The 1853 bp EcoRI, XbaI fragment of p511 was cloned between the EcoRI and XbaI sites of pBIOS4 (a derivative of pBI121), replacing the 35S CaMV promoter of this plasmid, forming plasmid pBIOS177.

To construct pBIOS4, pBI121 was digested with EcoRI, the ends filled using Klenow polymerase and then religated forming pBIOS5. This plasmid was digested with HindIII, the ends filled using Klenow and an EcoRI linker ligated into the destroyed HindIII site, forming pBIOS4.

pBIOS176 and pBIOS177 were transformed into *Arabidopsis* and tobacco. The larger promoter fragment is predicted to contain the entire Ms41-A promoter region since the EcoRI site lies with the 3' end of the W11 transcript.

*Arabidopsis* results:-

a) Short promoter:- Histochemical staining reveals that GUS activity is observed in most tissues and is especially high in callus, (strong blue staining is detectable after a few hours in X-GLUC (5-bromo-4-chloro-3-indolyl glucuronide)).

b) Long promoter:- GUS activity was seen in callus, but no obvious blue staining was observed in the vegetative parts of primary transformants. However 75% of the 40 transformants had significant GUS activity in anthers. In the floral buds observed, GUS expression is detected just after the breakdown of the callose wall (floral stage 10); expression appears to be located initially in the tapetum and subsequently in the microspores. GUS activity is still present in mature pollen. However it is possible

that there is also GUS activity in the microsporocytes and tetrad microspores since the GUS substrate may not penetrate the thick callose walls surrounding the microsporocytes and tetrads.

5

Similar staining experiments were done with plants containing the 3 tapetum-specific promoter fusions - TA29 (Koltunow *et al.*, *Plant Cell*, 2:1201-1224 (1990)), A6 (Hird *et al.*, *Plant Journal*, 4:1023-1033 (1993)) and A9 (Paul *et al.*, *Plant Molecular Biology*, 19:611-622 (1992)) and with the microspore/pollen promoter LAT 52 (Twell *et al.*, *Molecular and General Genetics*, 217:240-245 (1989)).

A9 is definitely the earliest and with the A6 promoter, 15 GUS is expressed when tetrads are visible; by contrast the TA 29 promoter gives expression at roughly at the same time as Ms41-A; the latter also shows earlier expression in microspores than LAT 52. In seedlings of 5 out of 7 transformed plants, very low levels of GUS 20 expression is detected in aerial parts.

Tobacco results:-

a) Short promoter:- GUS expression appears to be 25 constitutive.

b) Long promoter:- Results were similar to those observed in *Arabidopsis*, ie expression is largely confined to the tapetum, microspores and pollen of the anther. Very low 30 GUS expression was seen in the aerial parts of seedlings, however no expression was detected in callus.

It appears that expression from the long promoter matches that of the Ms41-A gene, with very low level

"consitutive" expression. Expression in the anther is much stronger than predicted by the abundance of Ms41-A transcript in floral parts indicating that the Ms41-A message may be very unstable. Higher level constitutive expression observed from the short promoter suggests that there a constitutive silencer is present in the upstream region of the promoter between positions -1635 to -900 bp. The conserved pattern of expression of the long promoter between tobacco and *Arabidopsis* suggests that the long promoter will be useful in male sterility systems in a wide range of plant species. Examples 3 and 4 below demonstrate the use of the long Ms41-A promoter in male sterility systems.

15 Example 3

Expression of Barnase from the Ms41-A promoter in Tobacco and Maize

20 The timing of expression of the Ms41-A promoter in the tapetum is similar to that seen from the tobacco TA29 promoter, thus fusion to cytotoxins such as Diphtera toxin A (Thorsness et al., *Developmental Biology*, 143: 173-184 (1991)) and Barnase (Mariani et al., *Nature*, 347: 25 737-741 (1990)) will ablate the anther tapetum leading to complete male sterility. Thus the long Ms41-A promoter is linked to Barnase. A 1kb XbaI, HindIII (filled) fragment encoding Barnase is excised from pWP127 (Paul et al., *supra*) and cloned between the XbaI and SstI (filled) 30 sites of pBIOS177 forming pBIOS 177-Barnase (Figure 7).

This plasmid is used to regenerate tobacco and Maize transformants that are male sterile. Although the weak "consitutive" expression of the Ms41-A promoter should

prevent recovery of such plants, it is likely that these plants have reduced Ms41-A promoter expression. Thus no significant expression of Barnase occurs in vegetative tissues whereas expression is sufficient to cause tapetal cell death and male sterility.

Example 4

10 Expression of antisense Ms41-A from the Ms41-A promoter in *Arabidopsis*

The Ms41-A promoter can be used to downregulate the expression of genes essential for tapetal function thus causing complete male sterility. Downregulation can be 15 achieved by expression from the Ms41-A promoter of antisense or sense fragments of the target gene or by expression of ribozymes which will cleave the target gene transcript. Such a target gene is Ms41-A. To construct an Ms41-A promoter- Ms41-A antisense chimeric gene, RT-PCR 20 is used to generate a 1923 bp Ms41-A fragment from young *Arabidopsis* floral buds mRNA. The primers used are:-  
W3 Bam, 5' CGGATCCTTCTAAGACATCACG 3' (positions 54-75, Figure 3) and  
3'2, 5' AATGTACTACTACTACTACTTAGGAC 3' (positions 25 3001-2976, Figure 3).

This PCR fragment is cloned into pGEM-T forming p542, such that the 5' end of MS41-A is adjacent to the ApaI site of pGEM-T (Figure 7). The MS41-A SpeI, ApaI (filled 30 using T4 DNA polymerase) fragment is cloned between the XbaI and SstI (filled) sites of pBIOS177, thus replacing the GUS gene of pBIOS177 and forming pBIOS182 (Figure 8). This plasmid is used to transform *Arabidopsis*. A proportion of transformants are male sterile with a

phenotype that resembled that of the original Ms41-A mutant. Examples 5 and 7 below describe the use of the Ms41-A transcribed region in male sterility systems.

5       Example 5

Expression of a 35S CaMV promoter- Ms41-A antisense chimeric gene and a 35S CaMV promoter Ms41-A sense chimeric gene in *Arabidopsis*

10       As described in Example 4, downregulation of the Ms41-A gene by expression of Ms41-A antisense fragments, sense fragments or ribozymes, each driven from the Ms41-A promoter will lead to male sterility. However any 15 promoter that has the appropriate pattern of expression, ie is active in microsporocyte and/or tapetal cells of the anther at the time of Ms41-A expression, may be used to downregulate Ms41-A and cause male sterility. Thus a 20 CaMV 35S promoter is linked to an antisense Ms41-A fragment and to a sense Ms41-A fragment. The antisense construct is obtained by cloning the ApaI (filled), SpeI p542 MS41-A fragment between the XbaI and SstI (filled) sites of pBIOS4 forming pBIOS188 (Figure 8).

25       The sense construct is obtained by cloning the ApaI (filled), SstI p542 MS41-A fragment between the SmaI and SstI sites of pBIOS4 forming pBIOS186 (Figure 8). These 30 plasmids are transformed into *Arabidopsis*. A proportion of the antisense and sense transformants are male sterile with a phenotype similar to that of the original Ms41-A mutant plant.

Example 6Isolation of a Ms41-A orthologue from Maize

5     Most methods to use the coding region of the Ms41-A in a male sterility system require the isolation of the orthologous sequence either from the crop species of interest or from a close evolutionary relative. Such methods include antisense and sense suppression and the  
10    use of ribozymes. The degree of evolutionary conservation between orthologous protein sequences is variable and is probably dependant on constraints on protein function. Although orthologous protein sequences may be highly conserved, codon usage may be quite different, producing  
15    orthologous mRNA sequences that may have low homology. Thus, in order to downregulate the Maize version of Ms41-A, it is probably necessary to isolate the Maize version of Ms41-A. Given the *Arabidopsis* Ms41-A mRNA sequence, several approaches are possible for the  
20    isolation of the Maize orthologue. Some of which are outlined below:-

25    The Ms41-A cDNA can be used as a probe on a Maize Northern or Southern at low stringency to see if a mRNA or genomic band hybridises. This was unsuccessful indicating that these sequences are widely diverged. The  
30    *Arabidopsis* sequence can be used as a probe in more closely related species and the orthologues in turn used as further probes until the version in Maize is identified. The cloning and sequencing of such orthologues may also result in the identification of conserved areas that can be used in a degenerate PCR approach.

Antibodies to Ms41-A may also be useful since protein sequences and epitopes are generally more conserved than RNA/DNA sequences.

5 The approach used was to screen the Genebank and EST (Expressed Sequence Tag) databases for sequences that showed homology to the *Arabidopsis* Ms41-A DNA sequence. Four groups of sequences were identified according to the degree of sequence similarity. Alignments of these  
10 sequences are presented in Figure 9.

#### Group 1

15 This group contains the *Arabidopsis* Ms41-A cDNA and an EST sequence from rice OSS2204 (D40316) which was cloned from a shoot cDNA library (prepared from etiolated 8 day old seedlings).

#### Group 2

20 In this group are two pairs of almost identical *Arabidopsis* EST sequences (ATTS3975 (Z37232) and T43470) and (T21748 and R30405) which are presumably derived from the same transcripts and can be considered as two sequences. The R30405, T21748 and T43470 cDNAs were  
25 isolated from a library prepared using a mixture of RNA from various tissues. The ATTS3975 cDNA is from a library prepared from cell suspension culture. In addition, in this group is a rice cDNA isolated from a root cDNA library (seedling stage) OSR1187 (D24087).

#### 30 Group 3

In this group are 3 EST sequences and 1 cDNA sequence ATTS1074 (isolated from a cycling cells cDNA library). A partial EST sequence for ATTS 1074 is on the database (Z25611) and after identification of this sequence as

similar to Ms-41A the cDNA clone was obtained and the sequence completed. The other 3 sequences are all identical or almost identical to the ATTS1074 sequence.

5 The cDNA clones R65265 and T44526 were isolated from a mixed RNA library. ATTS2424 is a 3' sequence EST sequence from the same cDNA clone as ATTS1074, this clone (TAI231) was isolated from a cDNA library prepared from a cell suspension culture containing cycling cells.

10 Group 4  
This group contains sequences of 4 closely related plant transcription factors; Viviparous-1 from maize (McCarty et al., *Cell*, 66:895-905 (1991)) and rice (Hattori et 15 al., *Plant Molecular Biology*, 24:805-810 (1994)), ABI 3 from *Arabidopsis* (Giraudat et al., *Plant Cell*, 4: 1251-1261 (1992)) and a *Phaseolus vulgaris* embryo-specific acidic transcriptional activator PvAlf (Bobb et al., *Plant Journal* In press (1995)).

20 There is some amino-acid similarity between a region in the N-terminal of the Ms41-A protein and the proposed DNA binding domain of maize Viviparous-1. This region is highly conserved between the 4 transcription factors (>80 25 % amino-acid identity between all 4 sequences). This suggests that the Ms-41A protein may have DNA binding activity, although the MS41-A protein might be sorted via the ER, perhaps to be secreted, since Ms41-A has a putative signal peptide and 6 putative N glycosylation 30 sites.

The most closely related sequence to Ms41-A identified by this analysis is the rice OSS2204 sequence. This was obtained from the rice sequencing project and used to

probe a Maize cDNA library made in Lambda UniZap (Stratagene) from polyA+ RNA isolated from pre-meiotic to meiotic-stage male inflorescences. The cDNA isolated, Zm41-A, is approximately 2.2 kb in length and has a poly 5 A tail at its 3' end. Approximately 300 bp of 5' prime sequence is shown in Figure 10.

10 This sequence shows strong similarity to the rice OSS2204 cDNA sequence (84 % identity) but is only 53% identical to the *Arabidopsis* sequence. The ORF indicated underneath the DNA sequence is similar to both the proposed OSS2204 ORF (89 % identical, 94 % similar) and the *Arabidopsis* Ms41-A protein sequence (54 % identical, 65 % similar).

15 A dendrogram of the Ms41-A related sequences indicates that the Zm41-A sequence falls into group 1 (Figure 11). This indicates that this cDNA is a good candidate for the maize orthologue of the *Arabidopsis* MS41A gene.

20 Example 7

Expression of an actin promoter- Zm41-A antisense chimeric gene in Maize

25 The Zm41-A cDNA is linked in an antisense orientation to a rice actin promoter. The entire Zm41-A cDNA is excised from pBluescript SK- (Stratagene) as an XhoI (filled), PstI fragment and cloned into PstI, SmaI - cut pCOR113 (McElroy et al., *Molecular and General Genetics*, 231: 30 150-160). This plasmid is used to transform Maize by a particle bombardment technique. A proportion of the transformants are male-sterile with a phenotype similar to that of the *Arabidopsis* Ms41-A mutant. This suggests that the Zm41-A sequence is the functional orthologue of

Ms41-A and indicates that any sequence that falls within group 1 (Figure 11) is likely to encode a functional orthologue of Ms41-A.

5       Example 8

Molecular characterisation of Zm41-A gene(s)

a) Zm41-A gene transcription

BY RT-PCR this transcript has been shown to be abundant in anther RNA; in leaf and tassel RNA populations it is 10 detected at a lower level.

After comparison of the maize and *Arabidopsis* sequences it was thought that the cDNA was unlikely to be a full length clone. With the "Marathon cDNA amplification" kit 15 (Clontech, Palo Alto, CA, USA) 5'RACE experiments were conducted on mRNA extracted from maize anthers at the meiosis stage, which yielded additional 5' sequence. Two types of 5'RACE products were obtained and sequenced, the first contained approximately 150bp of additional 5' sequence as well as a 108bp insertion at position 244 in 20 the cDNA. The second RACE product contained approximately 130bp of additional 5' sequence. It is believed that the first RACE product may be the result of differential or incomplete splicing of the transcript resulting in a 36 amino acid insertion in the predicted peptide sequence as 25 well as the 52 additional amino acids at the N terminal of the protein. Even with these additional sequences the full length transcript is likely to be longer at the 5' end, based on comparison with the *Arabidopsis* protein and 30 the maize genomic sequence.

b) Isolation of and characterisation of maize genes which are orthologs to Ms41-A

The Zm41-A cDNA was used to screen two different maize

genomic lambda libraries. The first was a commercial library (Clontech, Palo Alto, CA, USA) elaborated with DNA fragments from maize line B73 plantlets. DNA was partially digested with *Mbo*I enzyme and the fragments 5 were cloned into the *Bam*HI site of EMBL-3 (Frischauf et al, *J.Mol.Biol.*, 170:827 (1983)). The insert DNA can be excised from the clone by the enzyme *Sal*I. The second was a lambda library kindly provided by R. Mache (Universite Joseph Fourier, URA 1178, Grenoble, France) elaborated 10 with DNA fragments from the Mo 17 maize line. DNA was partially digested with the enzyme *Mbo*I and the fragments were cloned into the *Bam*HI site of EMBL-4 (Frischauf et al, *supra*). The insert DNA was excised by the enzyme *Eco*RI. The genomic libraries screening was performed 15 following the instructions of Sambrook et al (Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratory Press, New York, 1989). 10<sup>6</sup> recombinant Lambda per library were screened and three rounds of screening were performed. Fourteen positive lambda clones were 20 isolated one of which was obtained from the library provided by R. Mache.

DNA from positive lambda clones was extracted and purified using Qiagen columns (Chatsworth, CA, USA) 25 according to the manufacturer's instructions. Then the clones were characterised by Southern analysis (*J.Mol.Biol.*, 98:503-517 (1975)) in order to establish classes. DNAs from the Clontech library were restricted with *Hind*III and *Eco*RI and double restricted with *Hind*III/*Sal*I. DNAs from the Mache library were restricted 30 with *Hind*III and *Eco*RI and double restricted with *Hind*III /*Eco*RI. DNA fragments were separated on agarose gel, denatured and blotted onto Hybond N<sup>+</sup> membrane (Amersham, Buckinghamshire, UK). The blots were hybridised with <sup>32</sup>P-

labelled Zm41-A cDNA isolated after digestion with *Bam*HI and *Xba*I (the resulting fragment is 2.1 kb long).

10 Ten lambda clones were different and were distributed in  
5 three classes:

class A comprising 5 clones (Z9, Z23, Z27, Z35 and Z36);  
class B comprising 4 clones (Z7, Z28, Z29 and Z33); and  
class c with only one clone, Z31, isolated from the R.  
Mache library.

15 In order to study the sequence of these three classes,  
the sub-cloning of three different genomic phages (Z31,  
Z33 and Z35) in the plasmid pBSII SK<sup>+</sup> (Stratagene,  
LaJolla, CA, USA) was performed according to the  
classical cloning method (Sambrook et al, *supra*).  
Hybridizing fragments were firstly selected. After the  
sequencing of the fragments' extremities with universal  
primers, oligonucleotides were designed and the  
sequencing was achieved using the walking primer method.

20 With the clone Z31, 7.8 kb of continuous sequence data  
were obtained (see figure 12). To determine the complete  
gene structure, we have sequenced the entire Zm41-A cDNA.  
This is 2109 bp in length and encodes a putative peptide  
25 of 587 amino acids. The comparison between the genomic  
sequence and the cDNA and 5'RACE sequences indicated that  
this gene contains at least 12 exons. The insertion  
reported in the longest RACE products corresponds to the  
end of intron 4. Thus, the two families of cDNAs might be  
30 explained by the presence of two splicing sites in this  
intron. In the genomic sequence upstream of the end of  
the RACE products, there was detected the continuation of  
the open reading frame of 270bp before an initiation  
codon at a *Nco*I restriction site. Assuming that this

initiation site is the right one, the length of the fragment which might contain the promoter sequence was 2.7kb from the *Hind*III site where the sequence starts to the *Nco*I site. Therefore the translation of the Zm41-A 5 Z31 gene should give a putative protein of 736 amino acids. The Z31 gene structure is depicted in figure 13.

With the addition of the unspliced sequence (homologous to the end of intron 4) a longer protein might be 10 obtained. Indeed, the longest open reading frame deduced from the genomic sequence Z31 including this insertion sequence exhibits two stop codons in frame. It is also worthwhile noting that there is a clear polymorphism here since the RACE products do not show these stop codons. 15 The mis-splicing phenomenon may be a regulatory mechanism for the expression of the the Zm41-A related proteins as has recently been demonstrated in maize for another gene (Burr et al, The Plant Cell, 8:1249-1259 (1996)). Therefore, either this gene codes for two proteins (736 20 aa and 131 aa) or it codes for the 736 aa and 772 aa proteins.

Moreover, a slight difference was observed between the Zm41-A cDNA and the Z31 genomic sequence in exon ten 25 where a small addition is present (15 bp replaced by 36 bp); this is also in agreement with genetic polymorphism between maize lines. The maize lines used to study the mRNA and the genomic sequence are divergent (A188, B73 and Mo17 respectively). In figure 14 there is provided 30 the alignment of the Z31 protein (736 aa) deduced from the longest open reading frame, with the protein deduced from the Zm41-A cDNA (587 aa). We found 15 amino acid changes as well as an additional 7 amino acids for the Z31 protein, these additional amino acids being located

at position 556 of the Zm41-A cDNA protein.

For the other two genes, Z33 and Z35, 2.9 Kb and 5.8 Kb were respectively sequenced (see figures 15 and 16). Z35 5 contains exon 3 in part and the complete exons 4, 5 and 6 from the Zm41-A cDNA. Z33 is similar to Z35 but it has a deletion of exon 4 and the 3' end of exon 3. the two have the insertion sequence found in the longest 5' RACE 10 products. In addition, the comparison of the Z33 and Z35 sequences indicates at two deletions in the Z33 gene with respect to the Z35 gene. The first one is 686bp long and starts in the 3' end of exon 3 and extends to the end of exon 4 (with reference to the Z31 gene structure). The latter is located upstream of the sequence homologous to 15 Z31 and the Zm41-A cDNA and is 808bp long (see figure 13). Moreover, these two genes differed in their 3' sequenced regions.

Due to the high level of conservation between these 3 20 sequences it is possible that the Z35 gene derived from Z31 via genetic rearrangements, deletions and/or insertions. Z33 has subsequent deletions from Z35.

Example 9

Genetic mapping of Zm41-A loci

25 58 single seed descent (SSD) maize lines derived from the cross A188 x HD7 (Murigneux et al, *Theor.Appl.Genet.*, 87:278-287 (1993)) were used for genetic mapping by RFLP technology. Hybridisation was performed with 30 radiolabelled Zm41-A cDNA (BamHI-XbaI fragment, 2.1 Kb) on blots containing DNA from SSD lines and parental lines, digested with HindIII or EcoRI. Linkage analysis with the other RFLP markers mapped on this population was done using the Mapmaker version 2.0 computer program for

Macintosh (Lander *et al*, *Genomics*, 1:174-181 (1987)) and map distances were calculated with Kosambi function.

Many polymorphic bands between parental lines were revealed: one or two major bands and a few faint bands. Three loci, named *Zm41-A.A*, *Zm41-A.B* and *Zm41-A.C* were found located on two different chromosomes. *Zm41-A.A* locus corresponding to major bands, was located on the long arm of chromosome 6 at 26 cM from the RFLP marker *umc132* and at 2 cM from the rflp marker *umc62* (Maize Genetics Cooperation Newsletters (MNL) (August 1995) 69:248). *Zm41-A.B* and *Zm41-A.C* loci, corresponding to faint bands were located on chromosome 2 and were separated from each other by 19 cM. The *Zm41-A.B* locus lies near the centomere between *umc131* (6 cM) and *umc055* (3 cM) markers (MNL, *supra*). The *Zm41-A.C* locus was on the longchromosomal arm between *umc055* (16 cM) and *umc022* (6 cM) (MNL, *supra*). According to the mutant maize genetic map, no obvious male sterile mutant is mapped in those regions. One dominant male sterile mutant, *Ms21*, discovered in 1950 has been assigned on chromosome 6 but not very precisely. This mutation gives sterility only in the presence of the *sks1* mutation. Interestingly, this mutation maps on chromosome 2, in the vicinity of the *Zm41-A.B*. Hybridisation on the blots containing DNA from SSD lines, with a *Z31* gene specific probe, demonstrated that the *Z31* gene corresponds to the *Zm41-A.A* locus on chromosome 6.

CLAIMS:

1. A recombinant or isolated nucleic acid sequence which:

5

a) encodes the Ms41-A protein from *Arabidopsis*;

b) encodes a Ms41-A like protein;

10

c) encodes the ms41-A protein from *Arabidopsis*;

d) encodes a ms41-A like protein;

15

e) comprises a promoter sequence which regulates expression of the Ms41-A protein from *Arabidopsis* or a promoter sequence which regulates expression of a Ms41-A like protein; or

20

f) hybridises under stringent conditions to Nucleic acid of a), b), c), d) or e) or would do so but for the degeneracy of the genetic code.

25

2. Nucleic acid as claimed in claim 1 a) wherein the DNA encodes a protein having an amino acid sequence as shown in figure 4.

3. Nucleic acid as claimed in claim 1 b) which includes the sequence shown in figures 12, 15 or 16.

30

4. Nucleic acid as claimed in claim 1 derived from the family *Brassicaceae* or Maize.

5. Nucleic acid as claimed in any one of claims 1 to 4 which comprises a promoter, a coding region and a

transcription termination region.

6. Nucleic acid as claimed in claim 5 having at least a part of the nucleotide sequence shown in figure 3.

5

7. Nucleic acid as claimed in claim 6 having the nucleotide sequence shown in figure 3 commencing with the base pair labelled 1.

10

8. Nucleic acid as claimed in claim 1 a), b), c) or d) which includes a promoter sequence which drives expression in a plant tissue involved in the control of fertility.

15

9. Nucleic acid as claimed in claim 8 wherein the promoter is a tapetum-specific promoter.

20

10. Nucleic acid as claimed in claim 9 wherein the promoter is the A3, A6 or A9 promoter derived from *Brassicaceae*.

11. Nucleic acid as claimed in claim 1 e) which is operatively coupled to a DNA sequence.

25

12. Nucleic acid as claimed in claim 11 wherein the DNA sequence encodes a disrupter molecule.

30

13. Nucleic acid as claimed in claim 12 wherein the disrupter molecule is a lytic enzyme, a ribonuclease, a protease or a lipase.

14. Nucleic acid as claimed in claim 13 wherein the disrupter molecule is a ribonuclease, preferably Barnase.

15. Antisense nucleic acid which includes a transcribable strand of DNA complementary to at least a part of a DNA molecule as defined in any one of claims 1 to 10.

5

16. Antisense nucleic acid as claimed in claim 15 wherein the antisense nucleic acid is under the control of a constitutive promoter.

10

17. Antisense nucleic acid as claimed in claim 16 wherein the constitutive promoter is the CaMV35S promoter.

15

18. Nucleic acid encoding a ribozyme capable of specific cleavage of RNA encoded by a DNA molecule as defined in any one of claims 1 to 10.

20

19. Nucleic acid as claimed in claim 18 which also includes an appropriate promoter sequence, eg a constitutive promoter.

20. Nucleic acid as claimed in any one of claims 1 to 19 comprising a 3'-transcription regulation signal.

25

21. Nucleic acid as claimed in any one of claims 1 to 20 which is in the form of a vector.

22. A host cell transformed with nucleic acid as claimed in any one of claims 1 to 21.

30

23. A process for preparing nucleic acid as claimed in any one of claims 1 to 22, the process comprising coupling together successive nucleotides, and/or ligating oligo- and/or poly-nucleotides.

24. A plant cell including nucleic acid as claimed in any one of claims 1 to 21.

5 25. A whole plant, or part of a plant, comprising cells as claimed in claim 24.

26. A protein encoded by nucleic acid as defined in claim 1 a), b), c) or d).

10 27. A protein as claimed in claim 26 which has the amino acid sequence shown in figure 4.

28. The use of nucleic acid as defined in any one of claims 1 to 21 in the preparation of a transgenic plant.

15 29. A method for the production of a transgenic plant which comprises the step of transforming plant propagating material with nucleic acid as defined in any one of claims 1 to 21.

**The transposed 35S Ac element is linked to  
the male sterile mutation 41-A**

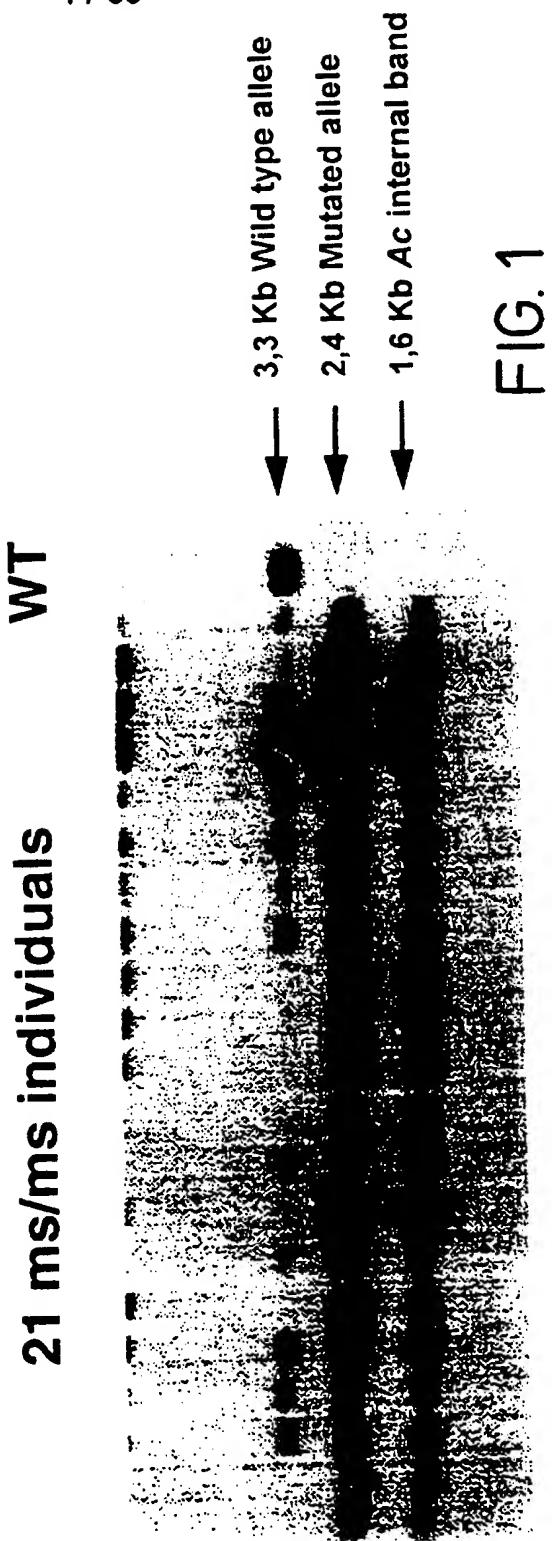
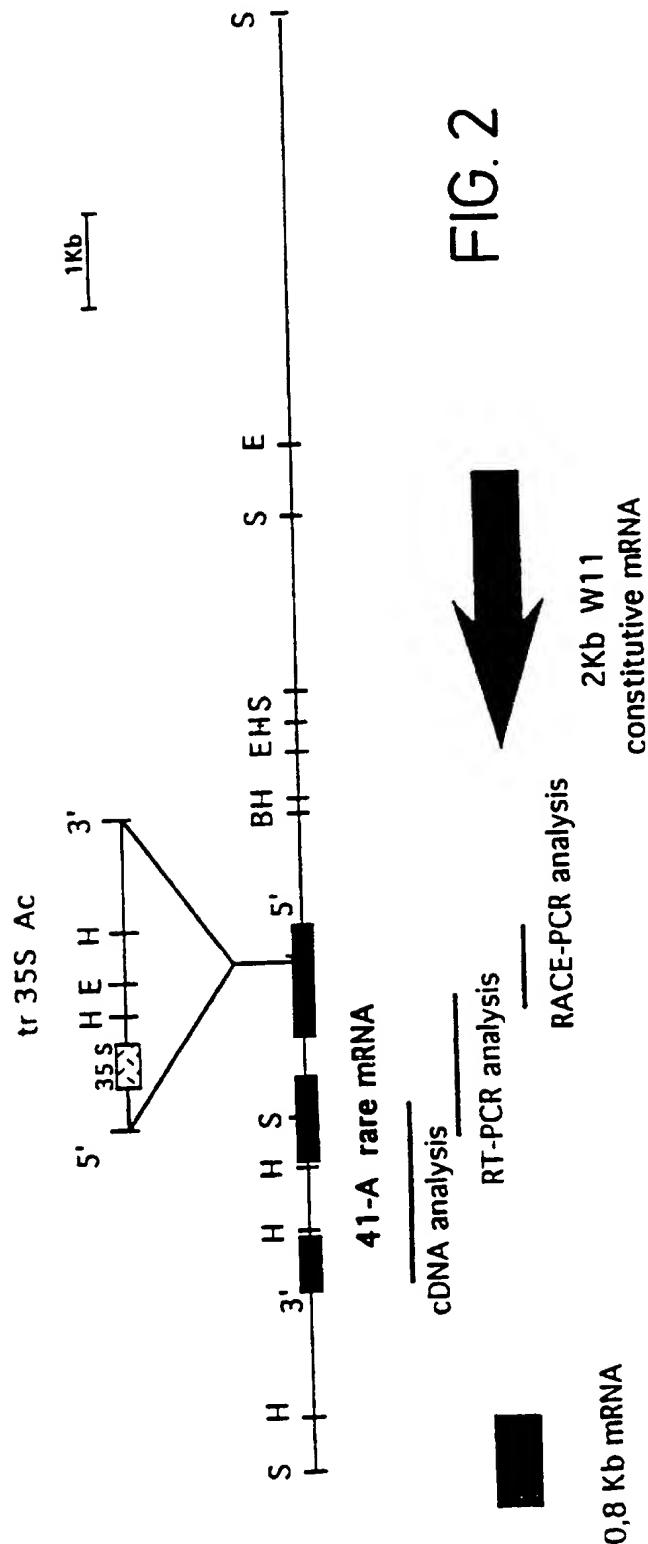


FIG. 1

2 / 35

### 41-A MALE STERILE MUTANT LOCUS



3 / 35

## FIG. 3(I)

EcoRI

TCCTAACCTTCTTGC~~GG~~CATTCTTATAAACTTCGTCA~~GG~~TTCAGAATTCTTAAATC  
 -1780 -1760 -1741  
 TTTTGCTGCTTCTTATAAAAGAAACATCATCTATTAAAGTTGTCTCGTTGGATTG  
 -1720 -1700 -1681  
 TTTTGATGACTTGGGAAATATTTATGTTAAGAACGGTTCA~~GG~~TATTGGTCATTGACTTTAT  
 -1660 -1640 -1621  
 ATATTATATCGTAACC~~GG~~ATGATGTGATAGTGGGCCTTAGATCAACAAACATGC~~GG~~AAAAACA  
 -1600 -1580 -1561  
 GAAGCAGAGGCCGTTCAACGGAGCATAATAAAATTGCATTCTCTGTCTTTGTTTAG  
 -1540 -1520 -1501  
 GTTTTTTTTTAACTGATAGATGTGCCGTGAAAATAATTTGATATTAAAAATT~~GG~~CACA  
 -1480 -1460 -1441  
 ACAAAACATTCTTA~~GG~~ACTGACCCACCCATCTATCTGCTATTCCCACGCCAAGGAAAATAA  
 -1420 -1400 -1381  
 TAATAATAGCGAAATTGATTTACATTTATTGATAGATAATTG~~GG~~TATTGTTAAG  
 -1360 -1340 -1321  
 ATTAACAGATTTAAGGGATTAAAGTGGAAAAGGTAAACCGAACACTTGCCATTAC  
 -1300 -1280 -1261  
 TGATTTACAACAATCCAAATTAAAAACAAATGGTCCCAGTTTTAGGGTTGTCACTTAA  
 -1240 -1220 -1201  
 ATTTATCGAAATATTTACACTTTAATTGGTAAACATAATGGACAGAAAAACAAATATT  
 -1180 -1160 -1141  
 HindIII  
 GTGACAAACAAAAAAACATGTTTCACCAAGAAAAACAAAAACAAAAAGATGTAAGCT  
 -1120 -1100 -1081  
 TTTCTTACATCTGTACAAAATAAAAGCAGACGAAATTGTACTTTATTTCCTTATTAAAT  
 -1060 -1040 -1021  
 TGTGGTATGTTTATATGTTGAAAAGTAGAATGGATAACCAATAAAATTACTGCA  
 -1000 -980 -961  
 HindIII  
 TCTTAATAAAAGTTGGTTCAACCGGTTAAAATGTATT~~GG~~TATTGTTAACAACTTAAAG  
 -940 -920 -901  
CTTTTTCGATTATCGAATTGCAACAAACAAATATTAACAGAAAAAGGAATCATGTA  
 -880 -860 -841  
 TCTATTTCAATATCCTGTTTTCTTCCATTGGATATTAGATCTTTCTGAATT  
 -820 -800 -781  
 ATCTTGTCTTAAATTAAACAGAAAAAGATTAAAGTAACAGACGCTTGCTAATGGCA  
 -760 -740 -721  
 ACCGCAACAAACAAAGATAATTGAAACCGATCCACTTGGAATTCTTGTGATTGTAGA  
 -700 -680 -661

4 / 35

## FIG. 3(II)

AAAATTGACAAATTGCTTTGTATAAAAACAAAAATGTACCGTAAAACACACACATAA  
 -640 -620 -601

AAAATAAAAAGTGATAATGACAAACAAATAAAGAGGTATTTTCTTTATCTACTAATGT  
 -580 -560 -541

GATTATAAAAAATCGACATTGAAAATTCAACACATTTTCGCCAAAACCTGAAAAT  
 -520 -500 -481

GGTCTTATTATAACATAAATTAGTTTTGTCTTCTATTATATATTCAATAACTCATC  
 -460 -440 -421

CCAACTTGAACAAACCTATAAGTTCCGTAGTGTCTTTCTGTGACAAAAAATACTA  
 -400 -380 -361

CCTAACGAGGGATAAGCACAAAAACATGATTATGTTCTCTAATCATTCTAAAAATCTA  
 -340 -320 -301

CAGGAATATTCCCTTTCAAGTTTCTTTCTAAATGCATTTCTTAGTTCTTCATAATT  
 -280 -260 -241

CAGTGAGTTTAATAACAATAATAAAAAAAAGAGCATCTTAATTGAACCTAAAAATAAT  
 -220 -200 -181

CGGAAGAAAAACCAAAAAGATAGAGAGTAAGATGCACGCCCTAAAGATCGAACGGTTAAT  
 -160 -140 -121

AGAATCAGGTTAGTGAAGAGAGATATTAAAAGTTGTTGCTGTGGCAAAAACATATAAT  
 -100 -80 -61

TTCCTCACACAAACAAAAAAATAAAATCAAACACAAAATCCCGTAGCATCGTAACAGT  
 -40 -20 -1

AATTGCTATTATCTCCTCACCCCTCCGTTCGTTCCCTCTGCCCCTTCAATTCC  
 1 20 40 60

TTCTAAGACATCACGGTCTCTCTATAAAAACAGTACCTACCTCTTCTTCTTCTTC  
 80 100 120

M S P P S A T A  
 ATTGCTGACTTCGTTACACTGAAAACAAATACCTATGTCACCGCCGTCGGCAACCGCC  
 140 160 180

G D I N H R E V D P T I W R A C A G A S  
 GGTGACATCAACCACCGTGAAGTAGACCCGACCGATCTGGCGCGCTTGCTGGAGCCTCC  
 200 220 240

C Q I P V L H S R V Y Y F P Q G H V E H  
 GTCCAGATCCCTGTCCTCACTCTAGGGTTACTACTTCCACAAGGTACGTTGAGCAC  
 260 280 300

3' 35S-Ac 5'  
 II  
 C C P L L S T L P S S T S P V P C I I T  
 TGTTGCCCTCTCCTCTCACTCTTCCACCTCGCCGGTTCCATGTATCATCACT  
 320 340 360

5 / 35

## FIG. 3(III)

S I Q L L A D P V T D E V F A H L I L Q  
 TCAATCCAGTTGCTCGCCGATCCGGTTACCGACGAGGTCTTGCTCACCTTATTCTTCAA  
 380 400 420  
  
 P I T Q Q Q F T P T N Y S R F G R F D G  
 CCGATCACGCAGCAGCAGTTACTCCGACTAATTATTACGATTCGGCAGATTGATGGC  
 440 460 480  
  
 D V D D N N K V T T F A K I L T P S D A  
 GATGTTGATGATAACAACAAGGTGACTACCTTCGCCAAAATTCTCACGCCCTCTGATGCT  
 500 520 540  
  
 N N G G G F S V P R F C A D S V F P L L  
 ACAATGGAGGTGGCTTCCTCGTTCTGCTGATTCCGTCTCCCTCTGCTT  
 560 580 600  
  
 N F Q I D P P V Q K L Y V T D I H G A V  
 AATTTTCAAATCGATCCACCGGTTAGAAGCTCTACGTCACTGATATCCATGGAGCTGTT  
 620 640 660  
  
 W D F R H I Y R G T P R R H L L T T G W  
 TGGGATTTCAAGGCATATCTATCGCGGTACACCGAGGCGTCACTGCTAACAAACGGGATGG  
 680 700 720  
  
 S K F V N S K K L I A G D S V V F M R K  
 ACTAAGTTGTCAATAGCAAGAAGCTCATCGCTGGAGATTGGTTGTGTTATGAGAAAA  
 740 760 780  
  
 S A D E M Y I G V R R T P I S S S D G G  
 TCTGCAGATGAGATGTACATCGGTAGCGAACTCCGATCTCAAGCAGCGACGGAGGA  
 800 820 840  
  
 S S Y Y G G D E Y N G Y Y S Q S S V A K  
 AGTAGCTATTACGGAGGAGATGAGTATAACGGTTACTACAGTCAGAGTAGCGTTGCCAAG  
 860 880 900  
  
 E D D G S P K K T F R R S G N G K L T A  
 GAAGATGATGGGAGTCCGAAGAAGACGTTAGGAGATCTGGGAATGCTAACGTTGACTGCT  
 920 940 960  
  
 E A V R S I N R A S Q G L P F E V V F Y  
 GAGGCTGTACGATCGATCAATAGACCGCTCTCAGGGATTACCGTTGAGGTGGTGTAT  
 980 1000 1020  
  
 P A A G W S E F V V R A E D V E S S M S  
 CCGGCTGCTGGATGGCTGAGTTGTTGTGACACGCTGAAGATGTTGAGTCTTCAATGTCT  
 1040 1060 1080  
  
 M Y W T P G T R V K M A M E T E D S S R  
 ATGTATTGGACTCCTGGACTCGAGTCAGAGATGGCTATGGAGACTGAAGATTCTTCGG  
 1100 1120 1140  
  
 I T W F Q G I V S S T Y Q E T G P W R G  
 ATCACATGGTTCAAGGCATGTTCTACTTATCAGGAGACCGGTCCATGGCGTGG  
 1160 1180 1200

6/35

## FIG. 3(IV)

S P N K Q L <---- intron 1

TCTCCATGGAAGCAGCTTCAGGTATATGATGTTTGAAATGGCTTGTCTTATC  
1220 1240 1260

TCTGTGATGTTGAGTTAATGGAACAATTAGAATCGATCTGTATCTGTGTGCAAGC  
1280 1300 1320

CTTTAAGATGATGTTAAGTCTCATCCTGGTTATTCAAATGTCATTGGGTTTTGAATGT  
1340 1360 1380

TGTTTGATTGCTGTGTTGTTGAAGCTAAATATTGAAACAGGATAAGTTAAgT  
1400 1420 1440

CATACGAAAATGAATGTTCTGTCTCAGATTCACTCTATAAGATGAAATTGAAACTGGA  
1460 1480 1500

AGATTTGGCTTAGTATTGTgTGTgTTGAGCCTCCGTGATGTAGAGTTGTTTCATTATCC  
1520 1540 1560

TTCTTGGCCACGCATGTACATTGTTGTTAAACTAGAGTTCTCTGATTAGTCTTA  
1580 1600 1620

TGAGATACTCCTTTTGCAATATATTCACTTCCTCTGATTAGTTCTTGTGTTTAA  
1640 1660 1680

----->Q I T W D E P E I L Q N V K R V N P  
CTTGCCTAGATCACATGGATGAACTGAGATTCTGAAAACGTGAAGAGGGTGAATCCA  
1700 1720 1740

W Q V E I A A H A T Q L H T P F P P A K  
TGGCAAGTGGAAATTCTGCACATGCAACTCAACTGCATACCCCTTCCCTCCAGCAAAG  
1760 1780 1800

R L K Y P Q P G G G F L S G D D G E I L  
AGGTTGAAGTATCCACAAACCCGGAGGAGGGTTCTTGAGTGGAGATGGAGAAATCCTT  
1820 1840 1860

Y P Q S G L S S A A A P D P S P S M F S  
TATCCTCAAAGTGGACTGTCTAGTCAGCAGCACCTGATCCAAGTCCTCTATGTTCTCG  
1880 1900 1920

Y S T F P A G M Q G A R Q Y D F G S F N  
TATTCTACATTCTGCTGGCATGCAGGGAGGCCAGGCAATATGATTTGGGTCTTCAAT  
1940 1960 1980

P T G F I G G N P P Q L F T N N F L S P  
CCAACCGGATTCAATTGGAGGAATCCTCCCCAGCTATTCAACCAATAACTCTTAAGTCG  
2000 2020 2040

L P D L G K V S T E M M N F G S P P S D  
CTTCCTGATTGGGAAAAGTGTCGACTGGAGATGATGAACCTTGGCAGTCCGCCATCAGAT  
2060 Sali 2080 2100

N L S P N S N T T N L S S G N D L V G N  
AACTTATCGCCTAATAGCAACACCAACTAATCTGTCTCTGGAAATGACCTGGTTGGAAAC  
2120 2140 2160

## FIG. 3(V)

7/35

R G P L S K K V N S I Q L F G K I I T V  
 CGAGGCCCCCTTCAAAGAAAGTTAACCGATTCAAGTTGGCAAGATCATTACCGTG  
 2180 2200 2220  
  
 E E H S E S G P A E S G L C E E D G S K  
 GAGGAGCATTCTGAGAGCGGTCTGCAGAGTCTGGCTGTGAGAGGATGCCAGCAAA  
 2240 2260 2280  
  
 E S S D N E T Q L S L S H A P P S V P K  
 GAGTCCAGCGACAATGAGACACAGTGTCTTATCACATGCTCCTCCAAGCGTGCCTAAA  
 2300 2320 2340  
  
 H S N S N A G S S S Q<----- intron 2  
 CATTCCAACAGCAACGCAGGTTCTAGCTCCAAAGGTATATTCCGATCTCTCAAGTACA  
 2360 2380 2400  
 HindIII  
 ATAATCAATTGAATCAGTTGCTATAAGCTTTTATTACTGTGACAAAGGAATTTC  
 2420 2440 2460  
  
 TTCCCTTCCCAGAACTATATTATGTAGAGTAGGAAACACAATCATGATTCTGATATGA  
 2480 2500 2520  
  
 CTTGACTGATGATGATACTTGTgAAAACTATCTATATCTCTTCAGTAATCAGTCGCCT  
 2540 2560 2580  
  
 TGAGCTAATGGAATTGGAACTTGAAACATTACTTGATTAACTTTCAATAGCATAA  
 2600 2620 2640  
  
 GCNTTCCTGTTCATCATATATGTTCACTATACTTGTATGCTTTATTACTGCTGATAT  
 2660 2680 2700  
  
 TTACTATTCCTGCTATTTTTGGTCTCGTTAACGGTATAAGGACACAGAATTGGCT  
 2720 2740 2760  
  
 CTTTTATCCATCAGAACTAGACATTACTGTACAAGTAGATGAAGAATTATGTGGTCCAT  
 2780 2800 2820  
 HindIII  
 TACAAATTTAATTGCAAGAAAGCTGAGCTGCTGCTTATAGACGATTATAATGTTGGAA  
 2840 2860 2880  
  
 HindIII -----> G \*  
 GATCCTGAAAGCTGGAAATGATTTGTACTTTCTTTGTTGTGTGTTGACAGGTTA  
 2900 2920 2940  
  
 AAAAGTGAAGAAGTGGTGGATCTTGCTGGAATCTCCAAGTCTAACAGTAGTAGTAG  
 2960 2980 3000  
  
 TACATTATATAATTCTGTTCTGCAATTGACTTTCTCTGGCTTTCTTGCCAC  
 3020 3040 3060  
  
 GTGACGATTCCGGTTTACTTTCTTCTTTTATCAATTCTCAGACACATTG  
 3080 3100 3120  
  
 ATGAACATCTCGCTCTCATCTAACGTTAACTATTTTATTGGGTAAATGTCTGGATT  
 3140 3160 3180  
  
 GTCTTACCTAAACATGTTAAAGACTGATGTTATGCAGAGTGAAAACAGTAAATAATT  
 3200 3220 3240

8 / 35

## FIG. 3(VI)

AATGCTTATTCAATCCCTATGCAATGGATCTCAACTAACGGGCCAACCAGAGACTT  
3260 3280 3300

TACTAACTGTCTTTGCTTTAGTTAATATTCTAATAAAATAAAAGACTGCCAATAATA  
3320 3340 3360

AAATCGGACCATTTTATTCTCATATAAAATAAAAGAAGCTCAAGGGAGGTCCCTCCTAC  
3380 3400 3420

ACTTTCTGACTCCTTATGTTCTGTGATTCAACGGATCAGCTATAGCAT  
3440 3460 3480

TTCCAATTGTCAGTAAGTTAGGGTTGGTTGGATTAGCTAATAGCTACCAATGAG  
3500 3520

9 / 35

Sequence Range: 1 to 584

50 \*  
MSPPSATAGD INHREVDPPTI WRACAGASVQ IPVLHSRVYY FPQGHVEHCC  
100 \*  
PLLSTLPSST SPVPCIITSI QLLADPVTDE VFAHLILQPI TQQQFTPTNY  
150 \*  
SRFGRPDGDV DDNNKVTTFA KILTPSDANN GGGFSVPRFC ADSVFPPLLNP  
200 \*  
QIDPPVQKLY VTDIHGAVWD PRHIYRGTPR RHLLTTGWSK PVNSKKLIAC  
250 \*  
DSVVPMRKSA DEMYIGVRRT PISSSDGGSS YYGGDEYNGY YSQSSVAKED  
300 \*  
DGSPKKTFRR SGNGKLTAEA VRSINRASQC LPFEVVVFYPA AGWSEPVVRA  
350 \*  
EDVESSMSMY WTPGTRVKMA METEDSSRIT WFQGIVSSTY QETGPWRGSP  
400 \*  
WKQLQITWDE PEILQNVKRV NPWQVEIAAH ATQLHTFFFFP AKRLKYPQPG  
450 \*  
GGPLSGDDGE ILYPQSGLSS AAPDPSPSM PSYSTFPAGM QGARQYDFGS  
500 \*  
PNPTGFIGGN PPQLPTNNFL SPLPDLGKVS T2MMNPGSPP SDNLSPNSNT  
550 \*  
TNLSSGNDLV GNRGPLSKKV NSIQLPGKII TVEEHSESQP AESGLCEEDG  
SKESSDNETQ LSLSHAPPV PKHSNSNAGS SSQG

FIG. 4

10 / 35  
**FIG. 5**

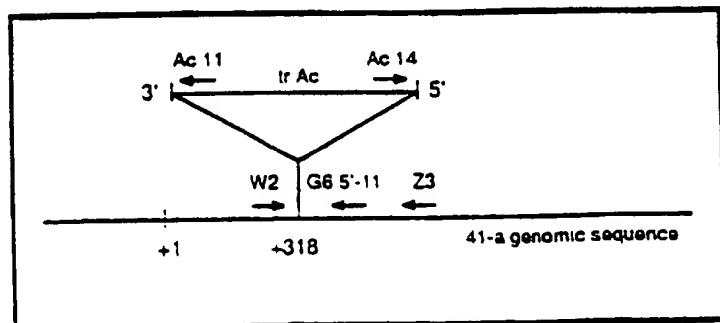
LIST OF PRIMERS

Ac 11	5' CGTATCGGTTTCGATTACCGTATT 3' located at position 4419-4443 on Ac sequence	25-mer
Ac 14	5' CGTTTCCGTTCCGTTACCGTTT 3' located at position 145-127 on Ac sequence	25-mer
W2	5' TGCTTGTGCTGGAGCC 3' located at position 221-237 on 41-a genomic sequence Concentration : 5296 ng / ul 1.0633 nmoles / ul	19-mer
Z3	5' GTTATCATCAACATGCCATCGAATCTGCCG 3' located at position 495-465 on 41-a genomic sequence Concentration : 13811 ng / ul 1.4555 nmoles / ul	31-mer
G6 5'-11	5' CTGCTGCTGCGTGATCGG 3' located at position 438-421 on 41-a sequence Concentration : 4943 ng / ul 0.8828 nmoles / ul	18-mer

COMMENTS :

Length of amplification product

W2 / G6 5'-11	217 bp
W2 / Ac 11	240 bp
Ac 14 / G6 5'-11	265 bp



11/35

Wild type Sequence	Cys Pro TGC CCT	Leu Leu Ser Thr Leu <u>CTC CTC TGT ACT CTT</u>	Alleles
Female A or B	TGC CCT <u>CTC CTC TC</u> (3' Ac 5') <u>C TCC TCT C</u> TA CTC TT	ms 41:: 35S Ac	
	TGC CCT <u>CTC CTC</u> AG <u>T CCT CTC</u> TAC TCT T	ms 41-1 (+7 bp)	
Male parent	TGC CCT <u>CTC CTC TC</u> (3' Ac 5') <u>C TCC TCT C</u> TA CTC TT	ms 41:: 35S Ac	
	TGC CCT <u>CTC CTC TC</u> T ACT CTT	Ms 41	
Revertant H	TGC CCT	<u>CTC CTC TC</u> T ACT CTT	Ms 41-R
	TGC CCT <u>CTC CT</u>	<u>C TCC TCT C</u> TA CTC TT	ms 41-2 (+5 bp)
Revertant K	TGC CCT	<u>CTC CTC TC</u> T ACT CTT	Ms 41-R
	TGC CCT <u>CTC C</u> AG <u>T CC TCT C</u> TA CTC TT	ms 41-3 (+5 bp)	
Revertant F	TGC CCT	<u>CTC CTC TC</u> T ACT CTT	Ms 41-R
	TGC CCT <u>CTC CTC T</u> G <u>T CCT CTC</u> TAC TCT T	ms 41-5 (+7 bp)	
Revertant C	TGC CCT	<u>CTC CTC TC</u> T ACT CTT	Ms 41-R
	TGC CCT <u>CTC CTC T</u> GA G <u>T C CTC TC</u> T ACT CTT	ms 41-4 (+9bp)	
Revertant M	TGC CCT <u>CTC</u>	<u>CTC CTC TC</u> T ACT CTT	Ms 41-1R (+3 bp)
	TGC CCT <u>CTC CTC</u> AG <u>T CCT CTC</u> T ACT CTT	ms 41-1 (+7 bp)	
Revertant A	TGC CCT <u>CTC</u>	<u>CTC CTC TC</u> T ACT CTT	Ms 41-1R (+3 bp)
	TGC CCT <u>CTC CTC TC</u> (3' Ac 5') <u>C TCC TCT C</u> TA CTC TT	ms 41:: 35S Ac	
Revertant L	TGC CCT <u>CTC CTC</u> G <u>T CCT CTC</u> TAC TCT T	Ms 41-2R (+6 bp)	
	TGC CCT <u>CTC CTC TC</u> (3' Ac 5') <u>C TCC TCT C</u> TA CTC TT	ms 41:: 35S Ac	

Footprints and alleles induced by 35SAC excision from the ms 41-a locus

FIG. 6

SUBSTITUTE SHEET (RULE 26)

12 / 35

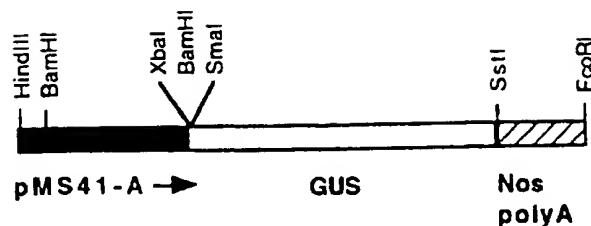
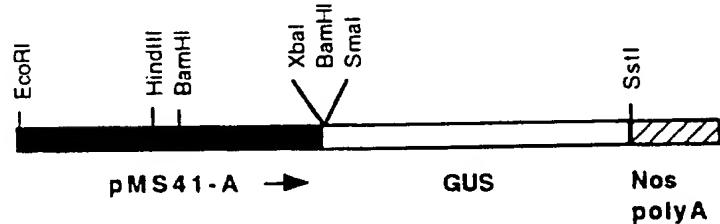
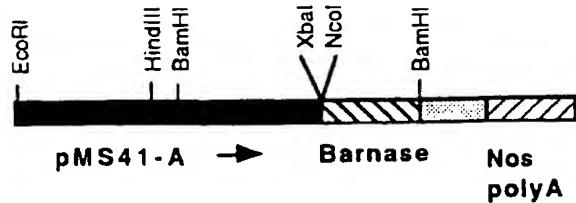
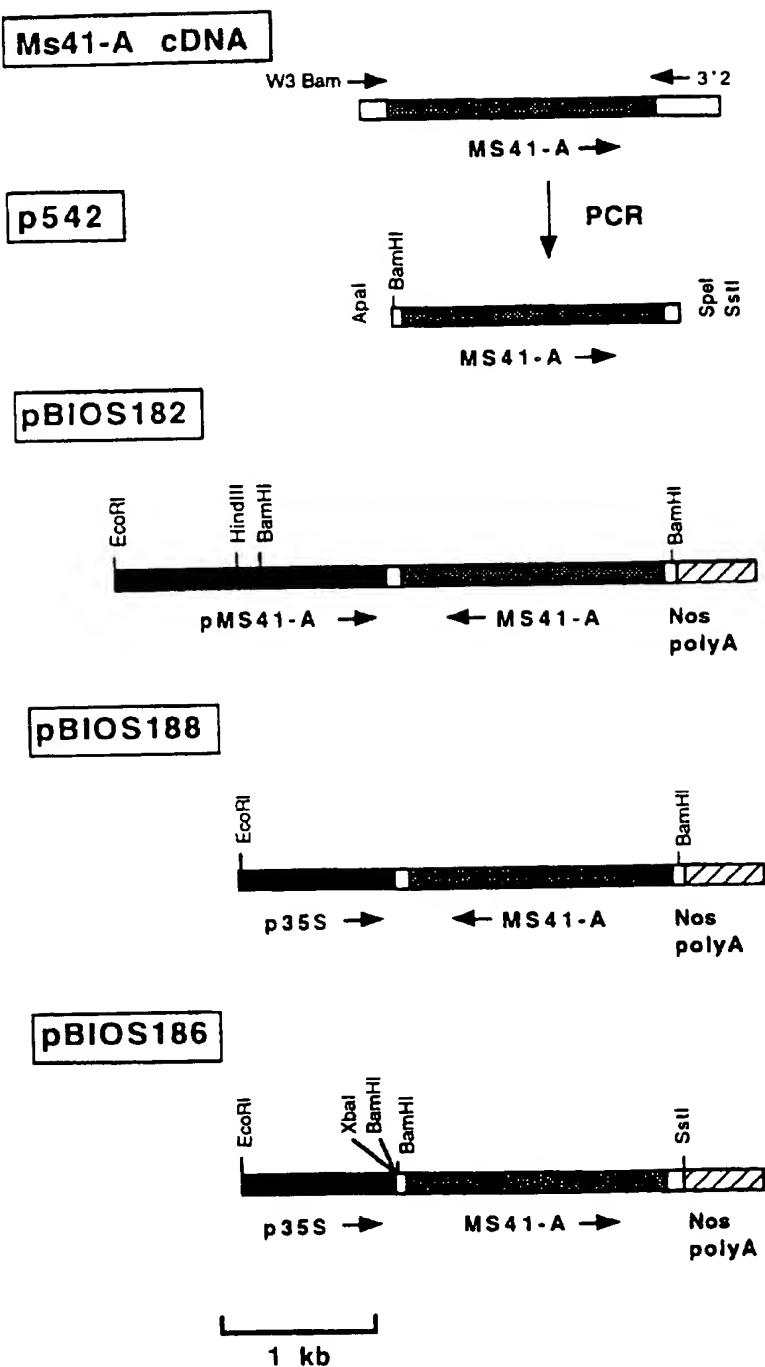
**pBIOS176****pBIOS177****pBIOS177-Barnase****FIG. 7**

FIG. 8



14 / 35

## Clustal Alignment of 41a related sequences

ZmVP1	515	.....	LLQKVLKQSDVGS
OSR1187	1	.....	EKRLTPSDV GK
ATTS3975	4	LRKHTYNEELEQSKRRRNNGNGNMTRLLTSGLSNDGVSTTGFRSAEALFEKAVTPSDV GK	
At41a	73	LADPVTDEVFAHLILQPIQQQFTPTNYSRFGRFDGVDDNNKVTI--FAKILTPSDANN	
OSS2204	1	..... AVKRLARI PHMFCKTLTASDTST	
Zm41a	1	DGSaedGVRKGTVKQRF SRM P HMF C K T L T A S D T S T	

ZmVP1	LGRIVLPKKEAEVHLP	---ELKTRDGISIPMEDIGTSRVWNMRYRFWPNNKSRMYLLEN
OSR1187	LNRLVIPKQXAERYFXLGGDGSX	-KQLLLXEDES-GKPKWRFYR SYWTSSQS--YVLXK
ATTS3975	LNRLVIPKHAEKHFPLPSNVSV	-KGVLNNFEDVN-GKVRFRYR SYWNSSQS--YVLTK
T21748	LNRLVIPKQHAEKHFPLPSPSAVTKGV	LNFEDVN-RKVRFRYR SYWNSSQS--YVLTK
At41a	GGGFSPVRCADSVPPLL	--NFQIDPPVQKLYVTDIHGAVWDFRHIYR--GTPRRHLLTT
OSS2204	HGGFSVPRRAAEDCFPPL	--DYSIQRPFQELVAKDLHGTEWRFRHIYR--GQPRRHLLTT
Zm41a	HGGFSVPRRAAEDCFPPL	--DYSQQRPSQELVAKDLHGTEWRFRHIYR--GQPRRHLLTT

ZmVP1	T-GEFVRSNE LQEGD	
ATTS3975	GWSRFVKEKNL RAGD	
T21748	GWSRFVKEKNL RAGN	
ATTS1074	1	..... GFSGFLR DDESTTTSKLM
At41a	GWSKFVN SKL IAGDSV VEMRKSAD E M Y I	-GVRRTPISSSDGGSSYYGGDEYNGYYQS S
OSS2204	GWSGFINKKKL VSGDCS A I P Q E V K M E N F D W G V R R A A	-QLKNAISF
Zm41a	GWSAFVN KKKL VSGD	.....

ATTS1074	MMKRNGNNDGNA	--AATGVRVR VEA VAE AVARAACQAFEV VVY PRASTP EFCVKAADV R
At41a	VAKEDDGSPKKTFR RSGN KLT AEAV-R S IN RAS QGL PFEV VY PAAGWSE FVVRAE DV E	.....

ATTS1074	SAMRIRWC S Q M R F K M A F E T E D S S R I S W F M G T V S A Q V A D P I R W P N S P W R L L Q V A W D E P D L	
At41a	S S M S M Y W T P G T R V K M A M E T E D S S R I T W F Q G I V S S T -Y Q E T G P W R G S P W K Q L Q I T W D E P E I	.....

At41a	LQNVKRVNPWQVEIAAHATQLH	-TPFPPAKRLKYPQP-----GGGFLSGDDG
ATTS1074	LQNVKRVSPWLVELVSNMPTIHLSPFSPRKIRIPQPFEFPFHGT K F P I F S P G F A N N G G	.....

At41a	E I L Y -----P Q S G L S S A A P D -----P S P S M F S -----Y S T F P A G -----M Q G A R Q Y D F G S F	
ATTS1074	E S M C Y L S N D N N N A P E G I Q G A R Q A Q Q L F G S P S P S L L S D L N L S S Y T G N N K L H S P A M F -L S S F	.....

At41a	N P T -----G F I G G N P P Q -----L F T N N F L S P L P D L G	
ATTS1074	N P R H H H Y Q A R D S E N S N N I C S L T M G N P A M V Q D K K K S V G S V K T H Q F V L F C Q P I L T E Q Q V M N	.....

At41a	K V S T E M M N F G S P P S D N L S P N S N T -----T N L S S G N -----D L V G N R G P L S K K V N S I Q L	
ATTS1074	R K R F L E E E A E E E K G L V A R G L T W N Y S L Q G L E T G H C K V F M E S E D V G R T L D L S V I G S Y Q E L	.....

At41a	F G K I -----I T V E E H S E -----S G P A E S G L C E E D G S K E S S D -----N E T Q L S L S H A P P S V P K	
ATTS1074	Y R K L A E M F H I E R S D L L T H V V Y R D A N G V I K R I G D E P F S D F M K A T K R L P I K M D I C G D N V R K	.....

At41a	H S N S N A G S S S Q G -----	
ATTS1074	T W I T G I R T G E N G I D A S T K T G P L S I F A	.....

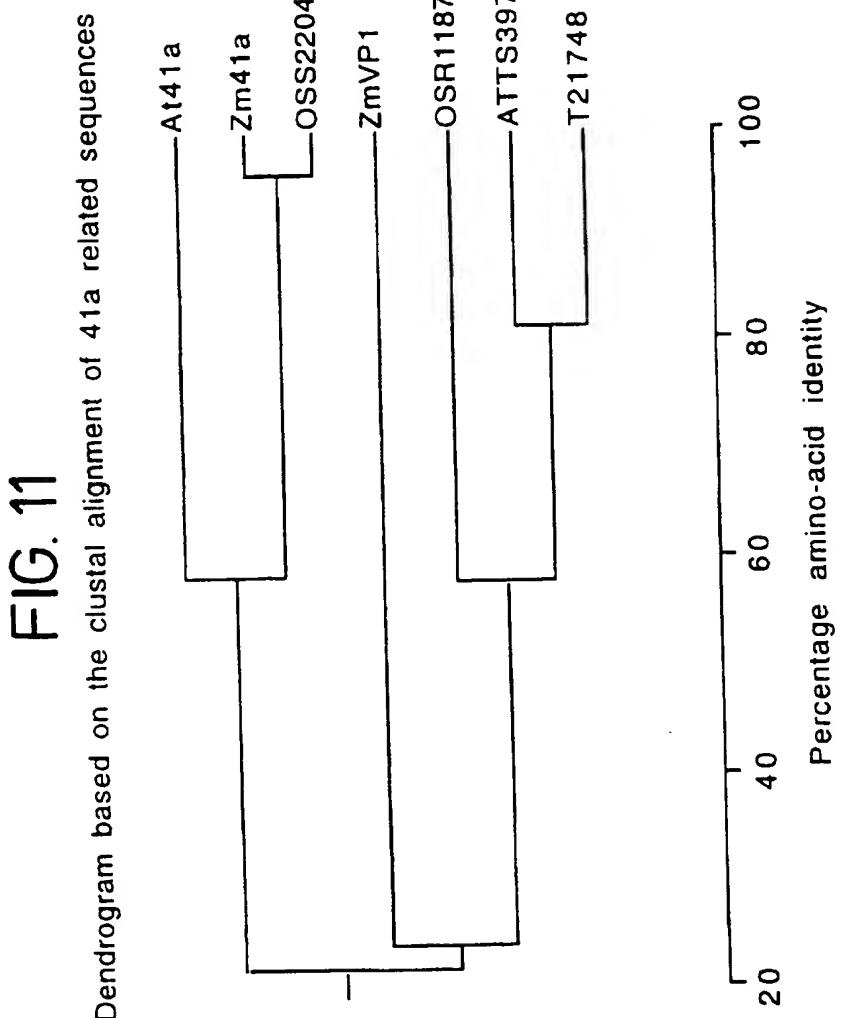
FIG. 9

15 / 35

## Zm41a 5 prime DNA sequence and proposed ORF

AGGACGGCAAGCGCCGAGGACGGCTACGGAAAGGGGAAACCGTGAAGCAGCGGTTTCGC 60  
 1 TCCCTGCCGTCGCGGCTCTGCCCATGCCCTTCCCCCTTGGCACTTCGTCGCCAAGAGCG  
     D G S A E D G V R K G E T V K Q R F S R  
  
 GGATGCCGCACATGTTCTGCAAGACGCTACGGCTCCGACACCAGCACGCACGGGGTT 120  
 61 CCTACGGCGTGTACAAGACGTTCTGCCAGTCCCGAGGCTGTGGTGGCTGCCAA  
     M P H M F C K T L T A S D T S T H G G F  
  
 TCTCCGTGCCGCCGCCGCCGCCAGGACTGCTTCCCCTCTGGACTACAGCCACCGC 180  
 121 AGAGGCACGGCGCGCGCGCTCCGTACGAAGGGCGGAGACCTGATGTCGGTGTGG  
     S V P R R A A E D C F P P L D Y S Q Q R  
  
 GACCGTCGCAAGGAGCTTGTGGCCAAGGATTGACCGAACCGAGTGGAGGTTCCGCCACA 240  
 181 CTGGCAGCGTCTCGAACACCGGTTCTAAACGTGCCTGGCTCACCTCCAAGGGGTGT  
     P S Q E L V A K D L H G T E W R F R H I  
  
 TTTATCGAGGGCAGCCCCGCAGACACCTTTAACCACTGGATGGAGTGCCTTGTCAACA 300  
 241 AAATAGCTCCCGTCGGGGCGTCGTGGAAAATTGGTGACCTACCTCACGGAAACAGTTGT  
     Y R G Q P R R H L L T T G W S A F V N K  
  
 AGAAGAAAGCTTGTCTCAGGGGAC 323  
 301 TCTTCTCGAACAGAGTCCCCCTG  
     K K L V S G D

FIG. 10



17/35

## FIG. 12(I)

aagctttagt gactagttagt agtgatttgt tggttttttt tgagcttttgc cgcttggatt gctttttttt  
140 \*  
ttctcattct ttcttgagat caatactcac ttgttaaccga ggcaagagac accaattgtg tggggcct  
210 \*  
tgccggtaag ttttggccgg ggttgatttg agaagagaaa gctcactcgg tccgagggac cgtttgaag  
280 \*  
aggaaaggaa ttgaaaaaga cccggccctt gtggccctt caatggggag taggtttgcg aaaaccgaac  
350 \*  
ctcggtaaaa caaatccgcg tgcacactt cttatctgtc tgcatggatttgc ttttcacccct ctctcgccga  
420 \*  
ctcgattata ttcttaacgc taacccgact tggtagttgtg attaacttttgc taaatttcag ttgcgcctta  
490 \*  
ttcacccccc tctatgcgac ttccagtagt tcatctatcc catgttttac ccctatttgc ttggatctga  
560 \*  
gctgattgcg acttagagac taaaactgtg aacttatgaa cctgtgaata aaatactaag taaaacttagt  
630 \*  
agtccgaatg tttgtgatag tcatacgca cccaaatcaa tataaaaatg gtttaaggcc aatttcctt  
700 \*  
cgcaaaagata tggaaatgtca taacccgtca atccttcatg taacaatggt cgtgcgttcc ctcaaccata  
770 \*  
caaaggacata tggccgcact gaaaaggcag acacacatag ttttacatat tttctacgct agcacaata  
840 \*  
cctcggttctc cactctgca ctcacgaaaa cagtaacaaa aacttcaaca acatactagg catattttct  
910 \*  
ctccaaactg gtctaaaaac tcttttcaaa ctcacttgcg gcaaggtaat cgggacatttta gcacccgaat  
980 \*  
ccctttccata aactccaaatc tacttgcata ggggtgaaa tcacgatagt taatgtgcata ggttaagggggt  
1050 \*  
atggcgcgtt gcatttagct ttgcgtggaa ttgcgtgcgtt ttcccatgac gccttcactc tcgaaaccaa

18 / 35

## FIG. 12(II)

tgtcacattt tgagatcttg gactttgtt ccaccaaggg attgcgccat gcagctccctc acccgcggtcc  
 1120  
 \*  
 ggacggtagc acacgagagg aaccatgaag ggcgcctcga catgcgggccc ttggatgggt cgacgaaaaa  
 1190  
 \*  
 ggtcttaggt tcgcggccta atgtcgacg acccgatgtctcgtacaagg tctatagaac ggttagggca  
 1260  
 \*  
 taagggcacc tagttcaaaa aactcaaaag ggcacaaccg agggtagggt tggcagcggg cgacgaagcg  
 1330  
 \*  
 aatatagccg cacgtgccac cacacaaaat gagggttaat ctgcgtatgc gcacccgtctca ggacccatca  
 1400  
 \*  
 tgcatgtagg atccatcttc gatgtcaatc acgatccctc atgctttac gacccctctcg acacgcctc  
 1470  
 \*  
 gtgcgtatgc tagaggacat tgtcgacgga atatccctc tttgcctgt gacttggatg attatacatt  
 1540  
 \*  
 tatggtaggt gttatggatg ttataaatgg atgtatatga atgtgtgtgt atctatgtgt tggatgaa  
 1610  
 \*  
 tataaataat tattttctaa ctggtaagaa tcatttctgg tgacttaggtt cagtcgataaa aatttagtat  
 1680  
 \*  
 gtcttaattt gtttattatgt ctatgaaaat tagttatattt tagtttattt atttcaaaa gttacagacc  
 1750  
 \*  
 gacgaaaact agactatcag tcacaactgg taagaaggaa caacgacaac agagatgcgc agttactggc  
 1820  
 \*  
 ttactgcagc aagctaccgt ttctgcggc cgtgtacatt gaagcacagg tgctgtctaca ctctacgtc  
 1890  
 \*  
 tcgagtc当地 tataaaaaata gactgttggg cacctattgt acccgatccctc ctgttctgc tcctgc当地  
 1960  
 \*  
 gtactgaatt ctgctgtgc tacactccctc tgtccgcata catccacgtc tctctctctc gccc当地  
 2030  
 \*  
 tgccgc当地 atcactgtgc gctgtcccg catcgccgc tctctttttt tttcaccctt tcccgccca

19 / 35

## FIG. 12(III)

2170

tcttctcttt ttacatctgc aacggcaggc cggctgcggc agcggcagcg gcagctgacc agtgaccgac  
 2240 \*  
 caccccccaca ccactccggc gccccaatcc tcccccttct tcttttac tactactact gtactgcacg  
 2310 \*  
 gtcgccaagc gccagaacgc agtggagaac gggggcagg actccaacaa gcgttgattt ctgcccggcac  
 2380 \*  
 gcacggcactg ggcacgggca cgggcacggg cgtcccccct cactcacgca ccctgcgtct tttccggctg  
 2450 \*  
 ccgctgctgg ctggctggct ctggctcaca gctacaggct acagtgaccg ccacgcaacc cacactgtct  
 2520 \*  
 ctgtccctgt ctccctctcc cctccctagct cttagctggat aggtgggctc tggggaggag gaggagggtt  
 2590 \*  
 gctaggttagt agctgcctat aggccctcggc ccccattcat ggccattacc acgatgtgtc accccaccac  
 2660 \*  
 accggccctct ccgatgtgc ctccctcatg ataaccctct ccctgggtgg tttttttgc cttgttgcgg  
 2730 \*  
 tgcagccctcc accccccaccc tcctcattaa tcacttgcta gtcctctgctt tccctccgg ctccccgtcc  
 2800 \*  
 cccttctatgt gtttgcggcc cccgcagcag ccATGGCGGG GATCGACCTC AACGACACCG TGGAGGGAGGA  
 2870 \*  
 CGAGGAGGAG CGGGAGCCCG GCAACGCCTG CTCCCAAGCAG AGCCGGACCA GCTCCGCGGC CACGTTCCCG  
 2940 \*  
 CCGCCGCCGC CGAACCCAGCC GAGGCCGAGC GCGCGGTGT GCCTCGAGCT GTGGCACGCC TGCGCCGGCC  
 3010 \*  
 CCGTCGCCGC GCTGCCGAGG AAAGGGAGCG TCGTGGTGTG CCTCCCGCAG GGACACATCG AGCACCTCGG  
 3080 \*  
 CGACGCCCGCG GCCGCCGGCG GAGGCCGCGCC GCGGCCCGTC GCCCTGCCGC CCCACGTCTT CTGCCGCCGTC  
 3150 \*  
 GTCGACGTCA CTCTCCATgt gcgccgcgg gttccctactc aatgcgtgcg tgtgtggatt gcccgtgcgg

SUBSTITUTE SHEET (RULE 26)

20 / 35

## FIG. 12(IV)

gtgtgcggct tccactgact ctgtccctct tgcgcgtt gcagGCGGAC GCGTCCACGG ACGAGGTGTA  
 3220 \*  
 CGCCCAGCTC GCCCTCGTCG CCGAGAACGA Ggtgcgcgca agccacagtg ctccacccgc attggattcg  
 3290 \*  
 gcttggttt ctccctgcgt ccacagagac gagatttggg ctgatttggt gtttcttgcg ggcgttgctt  
 3360 \*  
 cgtgcaggAT GTCGCGAGGC GGCTGCGCGG ACGGTCGGAG GACGGCAGCG CCGAGGACGG CGACGAAGGG  
 3430 \*  
 ▼  
 GAAACCGTGA AGCAGCGGTT CTCGCGGATG CCGCACATGT TCTGCAAGAC GCTCACGGCC TCCGACACCA  
 3500 \*  
 GCACGCACGG CGGCTTCTCC GTGCCACGCC GCGCCGCCGA GGACTGCTTC CCGCCTCTGg tacgcttgcg  
 3570 \*  
 ttggcttggaa aagcttccat cttttgggtg cccgggtgt gctctcaagt gcgattctga atcatctgt  
 3640 \*  
 cttggggcgt gcaggACTAC AGCCAGCAGC GACCGTCGCA GGAGCTTGTG GCCAAGGATT TGCACGGAAC  
 3780 \*  
 CGAGTGGAGG TTCCGCCACA TTTATCGAGG tacatgaaca aataatgaga tacaagacga gcacatctac  
 3850 \*  
 ctatttcttt agcaaactta tgtgcgttgcgtt cgcctgaat cattcagtgt cagcgaatga tgtcaatggc  
 3920 \*  
 tgcaacttcag ttggtgattt ttacaggat ttgcattact tgtttggatt gagcaacttgg  
 3990 \*  
 gaatgcgttca tcttgcgttca cttaaagtccca ggatttgaag tcattgttca gtcactcttt tgctatatat  
 4060 \*  
 gtcaccatta tgtgatcaga actactaatg gtttatgtt gagagagata tacaaactat gtcaatgtt  
 4130 \*  
 cctgcgttgcgtt gcatggcaaa cttgtgcgc tatgcgttgcgc atttctcatg tcattggattt gttattttag  
 4200 \*  
 tcgtacttaa aatttaccat ttgtccatg aaaaatcatc tgatttatG TTCAGGAGTT CTGGTCCCGT

21 / 35

TTTAAGGAAT GTAAAAGAAC AAACATGAGA AGCTATGTCA TGTGTGGTCC TTGGTTTCTG ATGAATCTGC  
4340 \*  
ATCTGAATGT GATGCAGGGC AGCCCCGAG ACACCTTTA ACCACTGGAT GGAGTGCCTT TGTCAACAAAG  
4410 \*  
AAGAAGCTTG TCTCAGGGGA CGCCGTACTA TTTTGAGgt aggccacagc taacattgga gataattatc  
4480 \*  
acatgttgtt gttggccctt tctgaagatt cctcataatt ttcagGGGTG ATAATGGGGA GCTAAGACTT  
4550 \*  
GGAGTGCGCC GTGCAGCTCA GCTTAAAAT GGATCTGCTT TTCCAGCTCT TTATAACCAG TGCTTAAATC  
4620 \*  
TTGGTTCACT ACCTAATGTT GCACATGCTG TGGCCACCAA AAGTGTGTTT CACATCTACT ACAACCCAG  
4690 \*  
gtgatgatga atatacggt ttcactttaa tgctttgca tttcaattt ttcatgttgtt tggactctt  
4760 \*  
ttagatgatg tgaactgaaa tgtgcttata actatactct ttcaatttgc ggcgatttga aattgtgtca  
4830 \*  
tttttgttga tatcatttcc tgagttgtt cgaactatgt aattcatgtat tcttactgca attcaacatt  
4900 \*  
aagtgtatata taattacttt ttgaatttgc attgtactt acatttggac cttcaatata aatatagttc  
4970 \*  
cacagctctt ttttagata tcatgacaag tacgcaagta gatcttttgtt cccttatgtt tctcatgtgc  
5040 \*  
attttaccc ttttggaccc tgatgtgtt cttcaatgtt taccttttta ccaccaaca atgatggccc  
5110 \*  
tgatggcaat tattgttttcc caaaaatctt acagATTAAG CCAATCTGAA TTCATTATAC CATTTCGAA  
5180 \*  
GTTTATCAAG AGCTTCAGTC AACCATTTTC TGCTGGTTCG AGGTTCAAAG TGAAATATGA GAGTGATGAT  
5250 \*  
GCTTCTGAAA GAAAGtttgcgtgctacagt ttcataatcc tacatagatt tatgttgtt gacacatgag

22 / 35

## FIG. 12(VI)

agtattatgc agATGCACAG GGATCATAGC AGGAATTGGT GATGCTGACC CCATGTGGCG TGGTCGAAA  
 5320  
 \*  
 TGGAAATGTT TGATGgtatg ttgcctttta agcttaatg attcacttcc tgtataactt ttcaagggtggt  
 5390  
 \*  
 aaatttgtgt tacatatgaa aataatccat gtagataca tgtagataat aacatgtttc tttatacaga  
 5460  
 \*  
 acactaggcg tgtgcatcat gtagctgccc ttgcacatcta tttgcactat ttgccttgcta ataaaccaat  
 5530  
 \*  
 aagcaatctt gcataatctat ccaataatac aatgcacaac aaatgttcaa aattgcaatt gagagccatc  
 5600  
 \*  
 tatgcacccc gtgcctccctg agctgtctct gtttgatgta caagtttaat tgtaatgaca cattttttt  
 5670  
 \*  
 gcatgtaaat agtttcctt ctccagagca cattcttga tgagccatcat cttagaggca tgttgtatct  
 5740  
 \*  
 ttatctaama gagactgcct tggccagcc tggttccctt gatcagggt ctaagtaaat aagttcaatt  
 5810  
 \*  
 cattttgggtt tcttattgcc ctgccttgcgttgcacatttgcgttgcata taatacccttc ttgacttgcgtt  
 5880  
 \*  
 aagccagttc taaattgccc caatcttaat cctcttgcgttgcacatttgcgttgcata ttttgcataat aaccaatgg  
 5950  
 \*  
 ttcattttttgcgttgcata ttttgcataat aaccaatgg  
 6020  
 \*  
 ttcatttttgcgttgcata ttttgcataat aaccaatgg  
 6090  
 \*  
 ATTGAGCTGA CTAGTCAGT TTCAGGATCT CACATGTCTG CACCAAATGC AAAGAGACTG AAACCATGTC  
 6160  
 \*  
 TTCCCCATGT TAATCCAGAC TACCTAGTTC CAAgtatgcc ctgttctgcc cagatgttgc cttaatgatt  
 6230  
 \*  
 attttgcgttgcata ttttgcgttgcata ttttgcgttgcata ttttgcgttgcata ttttgcgttgcata  
 6300  
 \*  
 AATCTGCCCA ATTCCACAAG GTCTTGCAAG GTCAAGAATT ACTGGGTTAT AGAACTCATG ACAATGCTGC

23 / 35

## FIG. 12(VII)

TGTTGCAACT TCTCAGCCAT GCGAAGCAAC GAACATGCAG TACATTGATG AACGAAGTTG CTCCAACGAT  
 6370  
 \*  
 GCGAGTAACA TTATCCCGGG GGTTCCAAGA ATTGGTGTCA GAACACCACT CGGAAGCCCT AGGTTTCCT  
 6440  
 \*  
 ACCGTTGCTC AGGCTTGGG GAGTCTCAA GATTCCAAA GGTCTGCAA GGTCAAGAAG TATTTCATCC  
 6510  
 \*  
 CTACAGAGGA ACTCTGGTCG ATGCAAGCTT GAGTAATAGT GGCTTCCATC AGCAAGATGG TTCTCATGTG  
 6580  
 \*  
 CCTACTCAGG CCAGCAAGTG GCACGCACAG CTACATGGAT GTGCTTTCG TGGCAACAA GCACCAGCTG  
 6650  
 \*  
 TTCCATCTCA ATCCTCATCC CCACCATCTG TCCTGATGTT TCAACGAGGT GATCCAAAGA TGTCCCCATT  
 6720  
 \*  
 TGAATTGGG CATTCCACG TGAATAAGAA AGAGGATAGA CGCGCAATGT TTGTCCATGC TGGAGGCATC  
 6790  
 \*  
 GGAGGAAC TG AGCAAAACGAC GATGCTCCAG GCTCATCATG TTTCTGGAGG AACGGGAAAC AGAGATGTGA  
 6860  
 \*  
 CCGTTGAGAA ATCTCATCCC GCTGTTGCCG CTGCTTCAGA CAACAGGGAA GTTAGCAAAA ACAGTTGCAA  
 6930  
 \*  
 AATATTTGGC ATATCTTGA CCGAGAAGGT TCCAGCAATG AAAGAAAAGG GCTGTGGTGA CATCAACACC  
 7000  
 \*  
 AACTATCCAT CCCCCCTCCT GTCTTGAAG CAACAAGTGC CGAAATCGCT GGGCAACAGC TGTGCCACCG  
 7070  
 \*  
 tgagtgtctt acaccatgtt gaccccttga tgtctttctc gagtgaagta actcttaact attataaaaat  
 7140  
 \*  
 cctgcacGTT CATGAGCAGA GGCCTGTTGT TGCTAGGGTG ATTGACGTTT CAACAGTGGA TATGATGATC  
 7210  
 \*  
 TGATGTATTG GAAAATGTC CTGGAGgtga agtcatgcta gtaccacctc tgtcttcattg ctagtgacca  
 7280  
 \*  
 tgaacagcat caaagcattt taagctgact gttcttaagc acatcgctt ttgttgtgc cttgtttt  
 7350  
 \*

## FIG. 12(VIII)

7420  
\*  
tgcaGCTGT GTTGCCTAGT GTGGACAGTG TCGGTTTGAT GGTCGGTAT CGTGAAGACG GGATTTGATT  
7490  
\*  
GAGGATCTGG CCAGATTTGT ATCCTAGTTG TAGCTGTTAG AGCACTTGT ATGACAACCG TGAGTGCTCC  
7560  
\*  
GTGTTATCAG CACTAGTTGC TGTCACAAAC TTGCCTCTAT GTTCATAATC TGTATGCCAT GTCAGACCCA  
7630  
\*  
TTTATAGAGG GTTGTGTTGC TTGGCATACT TCTAGACTTA AAGCATTATT ATGAGAACAA ATTTGCTCTG  
7700  
\*  
Caccgtatct ttcttacttt caagttggca acggattaac ggtggaggag atgatctgag aggttagttg  
7770  
\*  
tgcgacgtat taatggtgtt acatataatta tgcttaggag cattctgcca gtcatttat catatacatg  
7810  
\*  
tcagcacttg atttgttaag tgttagttgtt agccttgcac tttgg

25 / 35

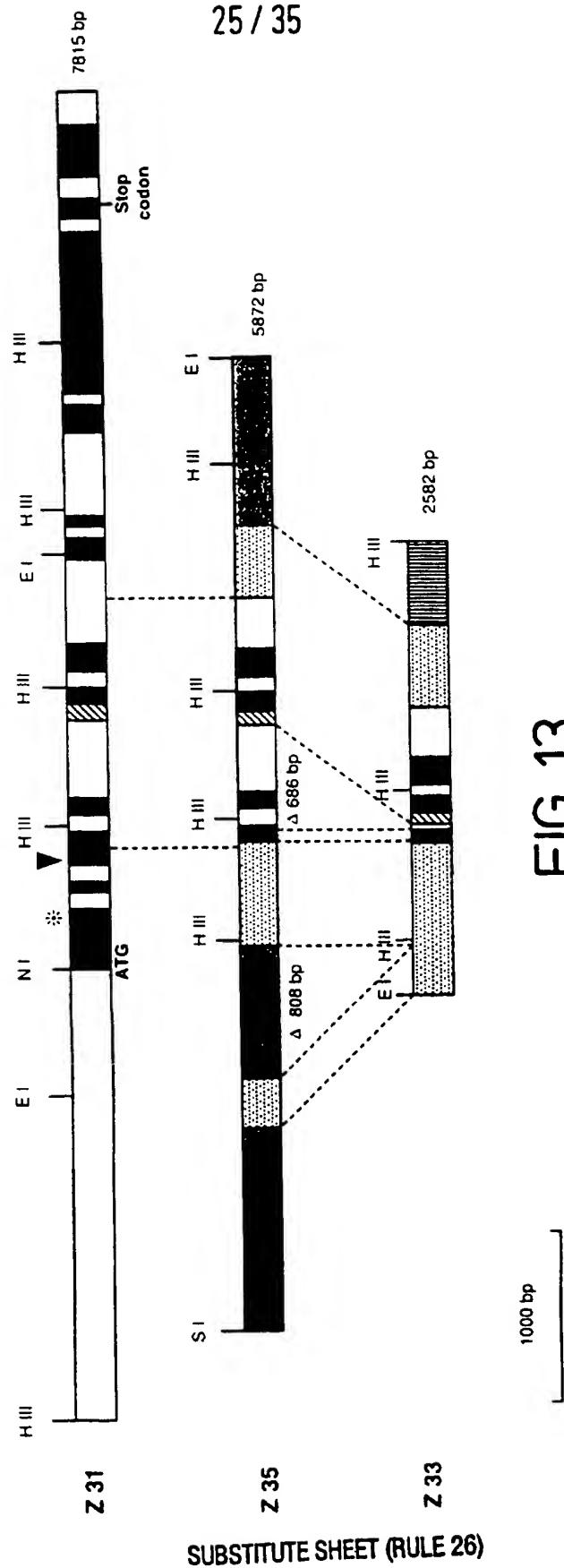


FIG. 13

SUBSTITUTE SHEET (RULE 26)

26 / 35

## FIG. 14

Z31	MAGIDLNDEEDEEEAEPGNACSQSRTSSAATFPPPPNQPRPSAAVC
Zm41-A	-----
Z31	LELWHACAGPVAPLPRKGSVVVYLPQGHIEHLDAAAAGGGAPPVALPP
Zm41-A	-----
Z31	HVFCRVVDTLHADASTDEVYAQALALVAENEDVARRLGRSEDGSAEDGD
Zm41-A	-----DGSAEDGD *****
Z31	EGETVKQRFSRMPHMFCYLTASDTSTHGGFSPRRAEDCFPLDYSQQ
Zm41-A	EGETVKQRFSRMPHMFCYLTASDTSTHGGFSPRRAEDCFPLDYSQQ *****
Z31	RPSQELVARDLHGTEWRFRHIYRGQPRRHLLTTGWSAFVNKKKLVSGDAV
Zm41-A	RPSQELVARDLHGTEWRFRHIYRGQPRRHLLTTGWSAFVNKKKLVSGDAV *****
Z31	LFLRGDNGELRLGVRRAAQLKNGSAFPALYNQCLNLGSLPNVAHAVATKS
Zm41-A	LFLRGDNGELRLGVRRAAQLKNGSAFPALYNQCSNLGSLPNVAHAVATKS *****
Z31	VFHIIYNPRLSQSEFIIPFSKFIKSFSQPFSAKSFRKVYEDDASERRC
Zm41-A	VFHIIYNPRLSQSEFIIPFSKFIKSFSQPFSAKSFRKVYEDDASERRC *****
Z31	TGIIAGIGDADPMWRGSWKCLMVRWDDDVDFRQPNRISPWEIELTSSVS
Zm41-A	TGIIAGIGDADPMWRGSWKCLMVRWDDDVDFRQPNRISPWEIELTSSVS *****
Z31	GSHMSAPNAKRLKPCPLHVNPDYLPNGSGRPDFAESAQFHKVLQGQELL
Zm41-A	GSHMSAPNAKRLKPCPLHVNPDYLPNGSGRPDFAESAQFHKVLQGQELL *****
Z31	GYRTHDNAAVATSQPCATNMQYIDERSCSNDASNIIPGVPRIGVRTPLG
Zm41-A	GYRTHDNAAVATSQPCATNMQYIDERSCSNDASNIIPGVPRIGVRTPLG *****
Z31	SPRFSYRCSGFGESPRFQKVLQGQEVFH PYRGTLVDASLSNSGFHQDGS
Zm41-A	SPRFSYRCSGFGESPRFQKVLQGQEVFH PYRGTLVDASLSNTGFHQDGS *****
Z31	HVPTQASKWHAQLHGCAFRGQQAPAVPSQSSPPSVLMFQGDPRMSPFE
Zm41-A	HVPTQASKWHAQLHGCAFRGQQAPAVPSQSSPPSVLMFQGDPRMSPFE *****
Z31	FGHFHVNKEDRRAMFVHAGGIGGTEQTTMLQAHVSGGTGNRDVTVEKS
Zm41-A	FGHFHVNKEDRRAMFVHAGGIGGTEQTTMLQAHVSGGTGNRDVTVEKS *****
Z31	HPAVAAAASDNEVSKNSCKIFGISLTERVPAMKEKGCGDINTNYPSPFLS
Zm41-A	HPAVATASDNEVSKNSCKIFGISLTERVPAMKEKGCGDINTNINTN-- *****
Z31	LKQQVPKS LGNSCATVHEQRPVVARVIDVSTVDMMI
Zm41-A	-----PKSLGNNSCATVHEQRPVVARVIDVSTVDMMI *****

27 / 35

FIG. 15(I)

28 / 35

FIG. 15 (II)

29 / 35

2170 \*  
atgcatatgt gatgtgaatt gagattcatt gagcaacaca aggattctgt gttggagatg gggtcttaat  
2240 \*  
atttctatca tgtaatatct tttggtagct tgcatcatat taataaaaata tctttggtgg cctcaggct  
2310 \*  
ggtggtaatg cttatgtgat tggtgattct gcaaaggctg agcagaagtg gcacgcctac tatgccacta  
2380 \*  
ctgagcaccc ctgaggagct tgggttact cttaacatgt gcatgactgg gctggacaag aagagagct  
2450 \*  
ctgtcttctt ctaggcttct gctgatggtt acacatcttgc tgcttaaggag atgaccaagc tctcaggat  
2520 \*  
ctcggacatt atcctataga cagagatctg cgactaattt gttaggttgg ttcttcataa ttttgtatag  
2580 \*  
gccttcctt ctcgctacat gaactaacta atgacagagg gtggaagtga cccatgaagc tt

FIG. 15(III)

30 / 35

FIG. 16(I)

gtcgacatgc aggtcaacgg atctattgaa ccagcagtct ttgcaattga gatttgactg ccggatttgg  
140  
\*  
tttcagcatg gatgcaccac cccacatcat gtggttctag agcatatagt ggtctttag cgcctaamag  
210  
\*  
ttttagtgc atcaaatgtc agaaaatataat cttcatctcc agaaaatatt agtacttcat aggatgaaaa  
280  
\*  
ttgttcaacc tgaaaataatt tatttcttgc atccttcagg ttgtatgcga aaccactaga ttgaataatt  
350  
\*  
caagaaaatct acagaggcag tcgtaaacaa ctatataatgc gcaagattga gcctaagggtt tgtagaccct  
420  
\*  
ttaattcata caagggcatt gccattttcc ccgttaatttc gatgcagctc cttagccat ataacaatga  
490  
\*  
aaaccaacga tcctgcaatc ctgaaaagggtt gaatttatgg gagaagcgta caactccctt agccaatgt  
560  
\*  
tccaaatgaag caccagcata caagaataag atagataaaat taacaggta taaaaatgtat actaataatcaca  
630  
\*  
tgttagtaaaa gaaacttaat cttccactg catcacgtat atgtgagtgc tccctggttt ttcattacag  
700  
\*  
tcttggatt tccattttat gtcgatgta ggtataggca tctgatggag gacgtttgc ctctactcccc  
770  
\*  
gcatgtgaag aaggacaacc aggacaaggt cgagtccaaag cagagcaagg gaaacacgct gaacaagttg  
840  
\*  
cttgaggatca ggagctgctt cagctgcctt tcttcgaggt atagatattc tactgtgcct ccacacagct  
910  
\*  
ggtggaaatt ttgttatcat agatacgtg cgggctgctt acatgtggga atcttacact gtataagtc  
980  
\*  
gtggcgcaaa tcaaataatctcc aacttgggtt tggccacact ttcgtgaaat gaatgttttc tgggctttca  
1050  
\*  
ggtatttgcgtt aaggagctcc cattttgcctc tggtgccaaa ttctctacta ggcaattgac gttttactgt

31 / 35

FIG. 16 (II)

catttgtgac atctgccttc ccacaattat aattgttcaa tatatgtatg cattagactt atcaatttt  
1190  
\*  
ttaacttatt gaattgtatg tgcataatg ttttcttcc atgtattaca ccacatgaca tagttcttta  
1260  
\*  
actaatggca gtgtacccctt ttaaccctt agatggctaa attcaaggga gaagatgatt tattagcagg  
1330  
\*  
ctctatgagc acagctgcac agtcaagaca taattcttgg gcctctgcag gtgattctca cccctacgct  
1400  
\*  
gacattgctt ggccttcaaa aatattcagt caagacaaaa agaacttact aatcaaatgt cattatcagt  
1470  
\*  
caatacttta agataagtag aatcgatgtc ccatacgaca ttctagccac gcacttaaac atgtgccaga  
1540  
\*  
tatgttcaga tcttgtgatt cagcagaccc tgacgcccag cgggcctccg cggaggcagt agccagatct  
1610  
\*  
ggccattttag tgccccgacg ccgctgctta ctcatccatc gcccgggtga cctgtcccccttccatc  
1680  
\*  
tctgtccatt gacaccaagc atgttcttcc ctgaactgtt ctaaaagtcc agtttcatgg ttgtttattc  
1750  
\*  
ttttgatcag gaaggagaga aagggagaat cagttagaag aaagaagagt ctgaaagctg agtaatttac  
1820  
\*  
ctcaacttta ctacccatgt tattaagatc tattgatgat cgtcccaactt actcctatga tgcacagact  
1890  
\*  
taatggatca tggactgaca tatttatcac gggtttggg ttgttttcttcc tcccaagtttt gttttaccag  
1960  
\*  
tggagacacg aagattggag gacataaggg cgcaacacag gactacagcg agggggaaagg ccagatcaag  
2030  
\*  
caggagacaa caagaggtgg gttgctgctc attcacaatt tgatatgtt gtttttcgt tgttataagct  
2100  
\*  
gaactocaca tgcagtttga aacatgttgt tactgtatgtg tttgtctatt acaggatgtg atagatggtg

32 / 35

## FIG. 16(III)

2170

atctctgtga gcagtatccc tccctcttag ctgatatgca gaggaagatt gctgatgagc tggacacaag

2240

tccgacgcct gcagcactgc ttggtgagga ttgccaagga ggaagactag aacaagcaag agcagcgtta

2310

atcagtgaca gagcatgatg ccatccagat gggacaagat aagtaagcag tcttatatacg tctgcccact

2380

cgagtttgt atatatttttta gtttgaagct tttgttttagt tcagtgttgc tatcgaaag ctaaaattat

2450

tttcttgcca ctcctctgc attgttttgc tgcttcagct cctggtgctt ctaatcgagt actatagaaaa

2520

gcatctctct tgataaaatcg ttgtgtgcaa atatagggtg cttatataat ccatcattag agtatgaggg

2590

gtgttttatt ctgtgtgctt cccacaaaaaa agagtagcct attataact ttgttattaga gcacatgacg

2660

ttcttaagtt tgaccacatt tctctactat tataatgcag ccataaagat tcaattttta tggggcac

2730

cataaagatg tttggcacca ttcttcccaa acatttatct actattataa tttgtgtgctt attcaatttt

2800

tagtattgtt aggggtgaag tcttagtctc aagatagcat attgttggttt gcctactccg acgactctga

2870

cgaggctgct gccccggcc aggagggagg tcaagaagcc taagaagccc aaggtaaga agcccaagGT

2940

GAAGCAACGA TTCTCGTGG A TGCCGCACAT GTTCTGCAAG ACGCTCATGG CCTCCGACAC CAGCATGCAC

3010

GTCGGCTTCT CTGTGCTGNG CCGCTCCGCC GAGGACTGCT TCCCGCTCT Agtacgcttg cgttggnttg

3080

gaaagttcc atcttttcgg tgccccggtg ctgctctcaa ggtgtgattc tgaatcatct gctcttgggg

3150

cgtgcagGAC TACAGCCAGC AGCGATCGTC GCAGGGAGCTT GTGGCCAAGG ATTTGCACGG AACCGAGTGG

SUBSTITUTE SHEET (RULE 26)

33/35

## FIG. 16(IV)

3220

AGGTTCCGCC ACATTTATCG AGgtacatga acaaatactg agatacaagc cgagcacatc tacctatttc  
 3290 \*  
 tttagcaaac ttatgtgctt gctcgccctg aatcattcag tgtcagcgaa tgatgtcaat ggctgcactt  
 3360 \*  
 cagttgatga ctgttagcgc ttttacagg atttgcatta cttgtttgga ttgagcactt aggaatgctt  
 3430 \*  
 catcttgct cacttaagtc caggattga agtcattgtt cagccactct tttgctataat atgtcaccat  
 3500 \*  
 tatgtgatca gaactaataaa tggttatatg tcgagagaga tatacaaact atgtcaatgt ttctgttgt  
 3570 \*  
 ctgcatttgc agccttgcgc gctatgctca gcatttctca tgtcattgggt tagttattgt agttgtactt  
 3640 \*  
 aaaaattacc attttgcata tgaaaaatca tctgattata tgtTCAGGAG TTCTGGTCCC GTTAAAGGA  
 3710 \*  
ATGTAAAAGA ACAAACATGA GAAGCTATGT CATGTGTTGCTTGGTAAATGATGAAATAT GCATCTGAAT  
 3780 \*  
GTGATGCAGG GCAGCCCCAC AGACACCTTT TAACCACTGG ATGGAGTGCC TTTGTCAACA AGAACCTTGT  
 3850 \*  
CTCAAGGGAC GCCGTACTAT TTTTGAGgtt ggccacaact aacattggag ataattatca catgttgggt  
 3920 \*  
ttggcccttt ctgaaggttc ctcgttaattt tcagGGGTGA TAATGGGGAG CTAAGACTTG GAGTGCAGCG  
 3990 \*  
TGCAGCTCAG CTTAAAATG GATCTGCTTT TCCAGCTCTT TATAACCAAGT GCTCAAATCT TGGTTCACTA  
 4060 \*  
CCTAAATGTTG CACATGCTGT GGCCACCAAA AGTGTGTTCC ACATCTACTA CAACCCCAAGg tgatgtgaa  
 4130 \*  
tatagcggtt tcacttaat gctttgcat gttcaattgt tcatgttgggtt ggcactcttt tagatgtatgt  
 4200 \*  
gaactgaaat gtgcttatta actactctt caattgacgg ggatttgaaa ttgtgtcatt gtgtgtgata

34 / 35

## FIG. 16(V)

tcatttcctg agttgttcg agctatgtaa ttcatgattc ttactgcaat tcaacattaa gtgatata  
 4270  
 \*  
 attactttt gaattgatat tgtcaattac atttggaccc ttcaatataa atcttccaa ttattgcct  
 4340  
 \*  
 ttttatccac tctttgtgt caagttctg caatttagaa gtatgcttc ttttgtatTT aattctttt  
 4410  
 \*  
 aggcacacaaa ttgttatttc ttcatgccat aatttctctg ttttattagt catagtaca aaaaatttt  
 4480  
 \*  
 tcaattgttg tggcgctag ctttgactgc tatggcggtg gcccggactgg cctgagatgg cgggtggccgg  
 4550  
 \*  
 atagcaccgc gagagcaacg tccagaggct agcagttcat tggtgttga gatttgtacc aatgattatc  
 4620  
 \*  
 tatattttaga gtttgttg gatacaccca tccattttat ccttggatTT cttttacaca gccatctaaa  
 4690  
 \*  
 ctctaaatTT agctaggatt ataaataacg tggatggatgc tcttaggtgg ctccctcaat ataggattag  
 4760  
 \*  
 tccatttttc tacagatggg gtgatagcat gcacattcta gcatacacat gcccggcc tggtaatgt  
 4830  
 \*  
 tggatttttt tctcacgcaa aagaatatac cggttcgTT aattatgtga tgcattttc tactttctg  
 4900  
 \*  
 ttttttagcc gatcatccga aggctaatga atattaccct gacccaaat tagtagcata tggatggcc  
 4970  
 \*  
 tatgcaccta tcctatcgTG gtatcactaa tccttctaaa tttgatatac tcttatctga ttcagcttg  
 5040  
 \*  
 tacttgattt aatttggctc cttgttaaca gtacggatgc tgcaaaaaat tccctgagga gaaaggTTga  
 5110  
 \*  
 aatctaaaaa ttgaaggcctc attggtccaa agcttacttc tatttgggg atgaggtgcg ttatTTacc  
 5180  
 \*  
 ttttctgcta tgcctgatt tcaggggaca ccagtgcaga tgcatgtagg gagaaacttg ttgcagttac  
 5250  
 \*

35 / 35

5320  
\*  
agaaaatggtt tccaaatatct actcttgcaa ttgaagatata ggagttactc cttgggttct ccttttagtt  
5390  
\*  
ttattatgct cgtccagtag acatgctccct gtagtaaact tatattcatg cttgtaattc catttacaat  
5460  
\*  
gtgaatattg tgtatagtag ccatgacatg ataatagatt gttagggtca ctcataaat attactatgt  
5530  
\*  
gccgtcacaa atatgggcac tccactaggg ttttagggttt tacctgttgt gcccagtttag ggtcactcat  
5600  
\*  
caaataattac agagggtatg ttccatttac agttggagta gatacgcattg acgggggcgc acatgagtt  
5670  
\*  
ttagtttgtt cgggatctca tgagtctgat tgacgtatgg cggatggctc tcgacgtgcg ggtcgacgac  
5740  
\*  
ggaacacttg cagcgcccat gttcggatgc agcgacagcc tccttgcgtc ttcaactcg cgacgagaga  
5810  
\*  
gagtggattt caggactgtc tgcttacagg agagaataa gctaatttct cagaatctta gaagctgatt  
5870  
\*  
ttacaacagg attgcttgct tacagagttt atcaactaaa aaagcgctat ggttcagaat tc

FIG. 16(VI)

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 96/03191

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 6	C12N15/29	C12N15/82	C12N15/11	C12N5/10	C07K14/415
A01H5/00					

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL Heidelberg, BRD AC D40316, 13 November 1994 SASAKI T. ET AL.: "Rice cDNA from shoot" XP002031170 see abstract &amp; UNPUBLISHED, SASAKI T. ET AL.: ---</p> <p style="text-align: center;">-/-</p>	1,6, 21-23

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- 'A' document defining the general state of the art which is not considered to be of particular relevance
- 'E' earlier document but published on or after the international filing date
- 'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- 'O' document referring to an oral disclosure, use, exhibition or other means
- 'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'a' document member of the same patent family

1

Date of the actual completion of the international search

Date of mailing of the international search report

20 May 1997

30.05.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patenttaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl  
Fax: (+ 31-70) 340-3016

Authorized officer

Kania, T

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 96/03191

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL Heidelberg, BRD AC Z34707, 25 June 1994 PARMENTIER Y. ET AL.: "The Arabidopsis thaliana transcribed genome: the GDR cDNA program" XP002031171 98 % identity to bp 98-1 (antisense) of Ms41-A see abstract & UNPUBLISHED, PARMENTIER Y. ET AL.: ---	15,21-23
A	TIBTECH, vol. 13, September 1995, pages 344-349, XP002031169 WILLIAMS M.: "Genetic engineering for pollination control" see the whole document ---	1-29
A	WO 92 13957 A (PLANT GENETIC SYSTEMS NV) 20 August 1992 see the whole document ---	1-29
A	WO 94 25593 A (CT VOOR PLANTENVEREDELINGS EN ;STIEKEMA WILLEM JOHANNES (NL); PERE) 10 November 1994 see the whole document -----	1-29

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/GB 96/03191

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9213957 A	20-08-92	AU 1202192 A CA 2103572 A EP 0570422 A US 5589610 A	07-09-92 08-08-92 24-11-93 31-12-96
WO 9425593 A	10-11-94	CA 2161515 A EP 0698098 A	10-11-94 28-02-96